GENOMIC INVESTIGATIONS OF THE ROLE OF DISINFECTANT-INDUCED ANTIBIOTIC RESISTANCE FOR PUBLIC HEALTH

A Dissertation
Presented to
The Academic Faculty

by

Minjae Kim

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
August 2019

GENOMIC INVESTIGATIONS OF THE ROLE OF DISINFECTANT-INDUCED ANTIBIOTIC RESISTANCE FOR PUBLIC HEALTH

Approved by:

Dr. Konstantinos T. Konstantinidis, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Frank Stewart
School of Biological Sciences
Georgia Institute of Technology

Dr. Spyros G. Pavlostathis
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Karen Levy
Rollin School of Public Health
Emory University

Dr. Joe Brown
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: May 01, 2019

ACKNOWLEDGEMENTS

I would like to express my sincere graditude to my advisor, Dr. Konstantinot T. Konstantinidis, for mentoring and supporting me throughout my graduate studies at Georgia Institute of Technology. I will be forever grateful for the opportunity to participate various research projects ranged from the engineered systems to human subjects, and for how he has shaped me into a hard worker and a more critical thinker. I would also like to thank my thesis committee members –Dr. Spyrod G. Pavlostathis, Dr. Joe Brown, Dr. Frank Stewart and Dr. Karen Levy –who have generously given their time and expertise to improve my work.

I am also gradeful to the members of Kostas' lab group for their support and assistance. Especially, I would like to thank Dr. Eric Johnston, Smruthi Karthikeyan, Dr. Janet K. Hatt, Dr. Luis Miguel Rodriguez-R, Dr. Despina Tsementzi, Carlos A. Ruiz, Dr. Angela V. Pena-Gonzalez, Dr. Maria Juliana Soto-Girón, Juan Camilo Camilo Castro Gordillo for their friendship and assistance during my thesis work.

Finally and most of all, I would like to thank my Parents, Mikyung Cho and Changsoo Kim, whose love and support was essential throughout this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Microorganisms occupy almost every habitat on earth such as soils, ocean, freshwater, and engineered systems (e.g., wastewater treatment systems), and are often associated with other multicellular organisms. Their communities play various important roles in controlling the biogeochemical cycles, and human and animal wellbeing (e.g., food fermentation, preventing infectious disease agents). Even though the great majority of microorganisms are beneficial to human and animal life and health, there have been rising public health concerns such as the recent emergence and spreading of antimicrobial resistant pathogens.

The advent of high-throughput sequencing technology and the accompanying development of bioinformatics tools for the analysis of the resulting sequence data in the past decade have enabled the high throughput characterization of the complex microbiota in various environments. Rapid analysis of microbial isolate genomes from various environments through whole-genome sequencing (WGS) has provided new insights into their ecological roles, identified novel species, and tracked the source of disease outbreaks (1, 2). For example, the identification of genomic islands (GIs) and single nucleotide polymorphisms (SNPs) among genomes of the same species can be used to identify the genomic determinants of antimicrobial resistance and to distinguish highly virulent from commensal strains of the species (3-5). Moreover, the rise of metagenomics (sequencing of the total microbial community DNA extracted from a target sample) has transformed our understanding of the microbial ecology and physiology of diverse

ecosystems, by bypassing the need to isolate the organisms in the lab, a major limitation of traditional lab-based approaches (2, 6, 7).

Although a substantial amount of work has been done in public and environmental health microbiology with omics techniques, there are still unresolved or debatable issues remaining. In this thesis, we combined traditional, culture-based laboratory techniques with cutting-edge, culture-independent omics tools to provide insight into several important research questions. Specifically, in **chapter 2**, we did metagenomic analysis of bioreactors, MinION-based long-read sequencing of microbial isolates, and molecular cloning to provide molecular evidence that exposure to the widely used disinfectant benzalkonium chlorides (BAC), a member of QAC family, can co-select for antibiotic resistant bacteria. These results contribute toward solving a high debatable issue in the literature for the past two decades, i.e., whether or not exposure to disinfectants promotes antibiotic resistance. In **chapter 3**, we assessed the effects of BAC-exposure of two different *Pseudomonas aeruginosa* strains, i.e., one pre-exposed to sub-inhibitory concentrations of BAC for three years vs. its non-exposed counterpart, to increasing concentrations of BAC for about two hundred generations (1-2 months) and applied transcriptomic analysis to reveal molecular mechanisms for the microbial BAC resistance at the whole cell level, including BAC-degradation by microbial enzymes. This work also identified microbial taxa and genes to BACs in non-target environments such as natural sediments and freshwater ecosystems. Finally, in **chapter 4**, we applied whole community DNA (metagenome) sequencing to 122 stool samples from young children in a low-income, urban neighborhood setting in Maputo, Mozambique, to assess the effect of shared, on-site sanitation intervention on bacterial gut infections. Our findings

suggested that the intervention induce small but significant differences in abundance of at least several microbial species, including *Veillonella parvula*, an opportunistic pathogen. However, pathogen load remain high in children with intervention compared to their matched controls, indicating alternative routes of infection that remain uncontrollable.

CHAPTER 1 INTRODUCTION

1.1 Microbial cross-resistance between disinfectants and antibiotics

The effects of antimicrobial resistance are already seen across the world. Excessive use of antibiotics and increased intercontinental travel nowadays have accelerated the spreading of antibiotic resistance pathogens worldwide, which leads to millions of casualties annually (8, 9). It is estimated that a continued rise in antimicrobial resistance by 2050 will lead to the death of 10 million people every year and a reduction of 2% to 3.5 % in Gross Domestic Product (GDP), which translates to up to 100 trillion USD (9). Even the proper and conservative use of antimicrobials contributes to the emergence of resistance, but widespread and excessive use aggravates the resistance problems. While the improper use of antibiotics is known to be the primary cause of the spreading of antibiotic resistance, excessive use of disinfectants such as quaternary ammonium compounds (QAC) can also contribute to the antibiotic resistance problem (10, 11). Antibiotics have a specific mode of action that typically differs from those of disinfectants such as QAC, which have non-specific and multiple target mode of actions. For this reason, QAC are thought to be functionally distinct from antibiotics, which implies the absence of cross-resistance between QAC and antibiotics.

Consistent with these interpretations, several studies reported the absence of cross-resistance between QAC and antibiotics in environmental, clinical, and industrial isolates. Further, some of these studies reported that disinfectant resistant pathogens showed decreased antibiotic resistance (12-16). For example, a previous study collected a total of 1238 isolates, including *Staphylococcus aureus*, *Klebsiella pneumonia*,

Escherichia coli, and *Enterococcus* sp. from both antibacterial users and nonusers, and provided evidence for the lack of cross-resistance between antibiotic and disinfectants (e.g., triclosan and QAC) in all tested bacteria (12). Also, both QAC resistant and non-resistant isolates of *Listeria monocytogenes* from an Iberian pig slaughterhouse and processing plant showed uniform susceptibility to 12 commonly prescribed antibiotics (15).

In contrast, several other studies have reported the presence of cross-resistance between disinfectants and antibiotics (4, 17-22). For example, sublethal concentrations of three widely used disinfectants, including QAC, caused *Salmonella enterica* serovar Typhimurium strain SL1344 to show fourfold-reduced susceptibility to ciprofloxacin, chloramphenicol, tetracycline, and ampicillin. Changes in protein expression such as ArcAB-TolC efflux pump and lipopolysaccharide O-antigens were implicated as the underlying causes for the decreased susceptibility (17). Another study reported that *L. monocytogenes* fed with 10 μg/ml of benzalkonium chlorides (BAC), a prominent member of the QAC family, showed increased minimum inhibitory concentrations (MICs) for ciprofloxacin, gentamicin, the dye ethidium bromide, and the chemotherapeutic drug tetraphenylphosphonium chloride (18). However, the issue of disinfectant-induced antibiotic resistance has remained highly debated because most, if not all, studies reporting the presence of the cross-resistance have not provided convincing evidence for the underlying molecular mechanisms for the reported linkage.

In chapter 2, using an integrated approach that combined metagenomics of natural microbial communities with gene cloning experiments and experimental

evolution assays of isolates, we show that the widely used BAC disinfectants can promote clinically-relevant antibiotic resistance.

1.2 Mechanisms of microbial BAC resistance

Benzalkonium chlorides (BAC) are the most commonly used members of quaternary ammonium compounds (QAC) disinfectants. BAC are cationic surfactant agents due to their amphiphilic property (charged nitrogen atom in the center of the molecule with long n-alkyl chain (C_8 to C_{18})). Thus, BAC have broad-spectrum biocidal activity (e.g., microbial, algal, fungal, and viral) and remain stable for both short and long-term usage (23). Accordingly, BAC are widely used as biocides in disinfectant solutions, food processing lines, domestic households, and healthcare facilities (24). Due to their wide use and mode of action, there has been rising concern that BAC may promote antibiotic resistance. Consistently, at least 40 outbreaks have been attributed to infection by disinfectant- and antibiotic-resistant pathogens such as *Pseudomonas aeruginosa* (25-35). Therefore, elucidating microbial BAC resistance mechanisms is important for monitoring and restricting the spreading of disinfectant-resistant pathogens.

The reported BAC resistance mechanisms include biodegradation by enzymes encoded in the genome of specific bacteria, overexpression of efflux pumps, reduction of the uptake or adsorption of BAC via downregulation of porins or cell surface modifications. Various kinds of efflux pumps have been associated with decreased susceptibility to BACs in phylogenetically diverse microorganisms. One such case is the multidrug resistance (MdfA) protein in *E. coli* (36). *E. coli* cells expressing MdfA are not only more resistant to a diverse group of cationic compounds including BAC, but also more resistant to clinically important antibiotics such as chloramphenicol and

erythromycin. QacA and QacB multidrug efflux pumps from the bacterial pathogen *Staphylococcus aureus* have been also shown to confer resistance to BAC and various toxic organic cations (37). Another example is the RND-type multidrug efflux pump MexAB-OprM. *mexAB-oprM*-deficient mutants of the plant pathogen *P. syringae* showed reduced tolerance to a broad range of antimicrobials such as BAC (38). Because the mode of actions of BAC involves the disruption of the cell membrane or cell wall (39), modifications of microbial cell surface have been hypothesized to be beneficial to BAC resistance. For example, BAC adapted *Salmonella enterica* serovar Enteritids showed a higher survival rate as biofilms compared to planktonic cells. Also, these BAC adapted cells showed greater cell surface roughness and a shift in cellular fatty acid compositions (e.g., phospholipids) (40).

The above-mentioned studies revealed several potentially important mechanisms and phenotypic characteristics of BAC resistant bacteria. However, the relative importance of these mechanisms for BAC resistance, at the whole-cell level, remain poorly understood. For instance, it is not clear whether or not all previously identified mechanisms are active/relevant for microbial communities that withstand high concentration of BACs or a few of them are quantitatively more important than others. Further, it is not clear if bacteria can biodegrade BAC by the production of catabolic enzymes. We previously described a microbial community degrading BACs under aerobic conditions, and we were able to isolate a key member (*Pseudomonas nitroreducens*) and identify its key biodegradation gene (i.e., amine oxidase) that is responsible for the BAC dealkylation, resulting in the 500 times reduction of BAC toxicity (41). Subsequently, another study identified a different biodegradation gene

cassette (42). Thus, it appears that BAC biodegradation is a major mechanism to tolerate high concentrations of BAC and the genetic markers to monitor biodegradation have recently become available. **In chapter 3**, using an integrated approach that combined genomics and transcriptomics with physiological characterizations of BAC-adapted isolates, we provided a comprehensive understanding of the BAC-resistance mechanisms in *Pseudomonas aeruginosa*, an important opportunistic pathogen of respiratory and other systems. Our findings also revealed potential genetic markers to detect and monitor the abundance of BAC-resistant pathogens across clinical or environmental settings.

1.3 Poor sanitation in undeveloped countries and its impacts on human health

The Joint Monitoring Program (JMP) funded by WHO and UNICEF estimated that only 68% of the world's populations live with improved sanitation ("basic" or "safely-managed" sanitation), providing the safe containment and separation of excreta from human contact to reduce public health risk (43). In contrast, 32% of humans without improved sanitation are exposed to environmental fecal contamination in the household or public settings, especially in urban areas of low-income countries with high population density (44-48). In low-income countries, urban infrastructure (i.e., improved sanitation) has not expanded to serve the needs of newly emerging populations. For instance, urban sanitation coverage has only increased from 37% to 47% in the poorest countries from 1990 to 2015 despite a much larger increase of the relative proportion of urban vs. rural populations in these countries (49). In sub-Saharan Africa (SSA), 921 million people still lack basic water and sanitation services (50), and, in particular, Mozambique showed one of the lowest coverage of sanitation services in the region (49).

The gut microbiota plays an important role in human health (51). For instance, the gut microbiota extracts nutrients from the food (52, 53), contributes to regular immune function (54), and prevents infections on enteric pathogens (55, 56). The composition of gut microbiota in early age is different from that of an adult. While one study reported that the age-affiliated gut microbiota changes from 3 days to 2 years after birth with major similarities between 2 years and adulthood (57), other studies have suggested that gut bacterial communities evolve into adult-like bacterial communities for up to 3 to 4 years after birth (53, 58) or even that gut microbiota is not yet established at 5 years of age (59). As the gut microbiota at an early age is more dynamic and variable compared to those of adults (60), it can also influence human host health in more profound ways; for example, via contributions to the maturation of the immune system in early days of life (61, 62). Furthermore, abnormal changes of the gut microbial community due to infection by enteric pathogens or prevalence/spreading of antibiotic resistance genes (ARGs) in the gut microbiome of infant/children could have long-term consequences for host health (63).

Unsafe Water, Sanitation, and Hygiene (WASH) conditions can facilitate the transmission of enteric pathogens among the residents through either direct contact or acquisition from the environment (64). The spreading of enteric pathogens often results in enteric infections that cause diarrheal diseases predominantly among the children and may contribute to chronic inflammation in the gut (65, 66). Children who experience chronic inflammation of their gut will suffer from serious health problems such as reduced absorption of nutrients (67, 68), environmental enteric dysfunction (EED) (69), growth faltering and cognitive deficits (70-72), stunting (73), and even death (74). While

sanitation interventions are generally believed to promote the reduction of enteric infections and the incidence of diarrhea in poor countries (75-80), other studies have shown no evidence of significant sanitation intervention effects on health (81-84). Insufficient coverage and use of improved latrines, difficulties in changing defecation behaviors, and a lack of proper waste management systems (e.g., sewerage systems), are thought to be the causes for the absence of effect. Furthermore, the majority of WASH studies have employed traditional protocols such as culture, microscopy, and PCR-related to detect the etiological agents (pathogens) of the disease. These methods, however, have known limitations (e.g., novel pathogens will be missed and pathogen loose cultivability when released in the environment) and could provide false-positive signal (48, 85-88), which represent another limitation in assessing the effects of sanitation interventions.

Two types of sanitation systems, an off-site and an on-site one, can be installed to collect and treat sewage in urban slums (89). Simplified sewerage, which is an off-site arrangement, differs from conventional sewerage because it is designed with reduced pipe diameters, gradients and depths, and non-compromised design standards such as placing sewers away from traffic loads generally under the sidewalks rather than under the center of the street to minimize excavation and the cost of pavement restoration (90). Simplified sewerage can be used in heavily populated urban slums lacking space for on-site sanitation (91), and consists of a sewer network and a wastewater treatment system (89). A flush toilet connected to a septic tank, which is an on-site arrangement, can also be used to collect and treat sewage. It only requires water to transport the waste from the pipes to a septic tank for proper storage and anaerobic treatment. Septic tank effluent can either be discharged in a soakaway (i.e., a deep hole used for drainage) before infiltration

in the ground or into a constructed wetland (89). This technology is suitable and widely used in urban slums where there is no off-site sewerage (92, 93).

In chapter 4, a controlled before-and-after (CBA) study of an urban sanitation intervention in Maputo, Mozambique is described. Specifically, compounds with ≤20 people were provided a single cabin "Shared Latrine," and compounds with >20 people received a "Communal Sanitation Block" (CSB) with one latrine cabin per 20 people residing in the compound (64). The CSB interventions also included facilities such as laundry and washing units, rainwater and municipal water storage, and handwashing stations. Controls were single-cabin pit latrines (i.e., open defecation) serving a minimum of 15 people and a maximum of 20 people so that the size of the control group was comparable to that of the sanitation intervention group. Also, the control group was enrolled during the recruitment of the sanitation intervention group to avoid having time gaps between the two study groups (64). The work described in the **chapter 4** employed metagenomic whole-community DNA sequencing to investigate whether or not sanitation interventions, especially a shared, on-site sanitation, reduced ea n enteric infections and the spreading of ARGs in young children as well as to describe the effects on the gut microbiome development, sidestepping the potential limitation of the traditional methods.

CHAPTER 2 WIDELY USED BENZALKONIUM CHLORIDE DISINFECTANTS CAN PROMOTE ANTIBIOTIC RESISTANCE

This chapter was published in:

Minjae Kim, Michael R. Weigand, Seungdae Oh, Janet K. Hatt, Raj Krishnan, Ulas Tezel, Spyros G. Pavlostathis, and Konstantinos T. Konstantinidis. *Appl. Environ. Microbiol.* 84.17 (2018): e01201-18.

2.1 Abstract

While misuse of antibiotics has clearly contributed to the emergence and proliferation of resistant bacterial pathogens with major health consequences, it remains less clear if the widespread use of disinfectants, a different class of biocides than antibiotics such as benzalkonium chlorides (BAC), has contributed to this problem. Here, we provide evidence that exposure to BAC co-selects for antibiotic-resistant bacteria, and describe the underlying genetic mechanisms. BAC-fed bioreactors inoculated with river sediment selected for several bacterial taxa, including the opportunistic pathogen *Pseudomonas aeruginosa*, that were more resistant to several antibiotics compared to their counterparts in a control (no BAC) bioreactor. Metagenomic analysis of the bioreactor microbial communities, confirmed by gene cloning experiments with the derived isolates, suggested that integrative and conjugative elements encoding a BAC efflux pump together with antibiotic resistance genes were responsible for these results. Further, exposure of the *P. aeruginosa* isolates to increasing concentrations of BAC selected for mutations in *pmrB* (polymyxin resistance) and physiological adaptations that contributed to higher tolerance to polymyxin B and other antibiotics. Physiological

adaptations included, for example, the overexpression of *mexCD-oprJ* multidrug efflux pump genes when BAC was added in the growth medium at sub-inhibitory concentrations. Collectively, our results demonstrated that disinfectants can promote antibiotic resistance via several mechanisms, and highlight the need to remediate (degrade) disinfectants in non-target environments to further restrain the spread of antibiotic resistant bacteria.

2.2 Introduction

While inappropriate prescription to humans and misuse of antibiotics in animal feed are thought to be the leading causes of the increased frequency of antibiotic resistance observed in recent years, there is an increasing concern that widely-used disinfectants such as quaternary ammonium compounds (QAC) have contributed to the antibiotic resistance problem (10, 11). However, this issue remains highly debatable because disinfectants typically have a different mode of action than antibiotics in order to avoid overlap between these two distinct classes of biocides that are used for different purposes. Further, previous studies have yet to elucidate the underlying mechanisms for the reported linkage between disinfectant-exposure and increased antibiotic resistance (4, 17-21). For instance, *Pseudomonas aeruginosa*, an important opportunistic human pathogen, acquired a 12-fold increased resistance to disinfectants upon disinfectant exposure, which was accompanied with a >200 fold increase in ciprofloxacin resistance (4). The increased ciprofloxacin resistance was attributed to specific mutations in the *gyrA* but it remained unclear if the mutations were indeed induced by the disinfectant, occurred spontaneously or as a result of the growth conditions. The different mode of action and target sites between QAC and ciprofloxacin, and the fact that QAC do not target the product of the *gyrA* gene argue in support of the latter interpretation.

Furthermore, other studies reported that antibiotic and disinfectant cross-resistance is absent in environmental, clinical, and industrial isolates, and disinfectant-resistant pathogens often show increased antibiotic susceptibility (12-16). Even the recent restriction in the use of triclosan and triclocarban, another family of widely used disinfectants in antimicrobial soaps, is not accompanied by consistent evidence on the linkage between triclosan exposure and antibiotic resistance (94). For instance, several studies reported decreased susceptibility of *Escherichia coli* to chloramphenicol, erythromycin, imipenem, tetracycline, trimethoprim, and other biocides as an effect of exposure to sub-lethal concentrations of triclosan (95). However, other studies suggested that there is no clear link between increased triclosan tolerance and increased antibiotic resistance, and actually found that *E. coli* with increased triclosan tolerance is more sensitive to aminoglycoside antibiotics (96).

Benzalkonium chlorides (BAC) are the most commonly used members of the QAC family of disinfectants, have broad-spectrum (i.e., bacterial, algal, fungal, and viral) biocidal activity, and remain stable for both short and long-term usage (23). Consequently, BAC are widely used as surface disinfecting agents in food processing lines (e.g., poultry facilities), dairy/agricultural settings, health care facilities, and domestic households, and are popular ingredients in over-the-counter cosmetics, hand sanitizers, and pharmaceuticals products (24). Therefore, BAC represent ideal molecules to study the effect of disinfectant exposure on microbial antibiotic resistance. Understanding whether or not exposure of bacteria to BAC may lead to increased antibiotic resistance and, if so, what the underlying molecular mechanisms might be, are important questions to answer for better regulating the usage of BAC and minimizing public risk.

To obtain insights into this issue, we exposed a microbial community originating from a river sediment inoculum (Calcasieu River, USA) to BAC for three years in three

aerobic fed-batch bioreactors: dextrin/peptone plus BAC (DPB bioreactor), dextrin/peptone (DP bioreactor), or BAC only (B bioreactor) as the main carbon and energy sources, as described previously (19, 97). This river is heavily contaminated with metals, polycyclic aromatic hydrocarbons and chlorinated/halogenated organic compounds. Therefore, its sediment represented an ideal inoculum for our purposes to identify organisms with resistance to BAC and other organics since BACs typically accumulate in non-target sediments and anoxic habitats (although the exact inoculum used in the bioreactors did not contain detectable concentrations of BAC at the time of sampling (97). Three years of BAC exposure led to significant changes in the composition of microbial communities compared to a control community (DP) under the same laboratory conditions (Figure S1A). The BAC exposure also selected for efficient degraders of BAC, which we reported previously together with the identification of a novel BAC-degrading gene cassette (98). Here we focused on the effects of BAC-exposure on antibiotic resistance. We also assessed the effects of BAC-exposure on shorter time scales by exposing isolates from the control (DP) and DPB bioreactors to BAC for about two hundred generations or less (1-2 months).

2.3 Materials and methods

2.3.1 Bioreactor development, and isolate characterization

All microbial communities (DP, DPB, and B) analyzed in this study originated from the same inoculum, from a sediment sample collected at the Bayou d'Inde, a tributary of the Calcasieu River, near Lake Charles, LA. The residual concentration of BAC in the sediment was less than the detection limit of the colorimetric method, i.e., 1 ug/g-sediment (99). Community development and bioreactor operation (e.g., substrates, feeding cycle, temperature) were described previously in detail (19). Briefly, the DP

microbial community was developed from the sediment inoculum with dextrin/peptone 50:50 (w/w) mixture as carbon and energy sources and maintained for one year. The DP community was subsequently used as an inoculum for DPB, which in turn was used as an inoculum for B microbial community. DPB and B communities were fed by dextrin/peptone plus BAC mixture and BAC mixture as carbon and energy sources, respectively. BAC mixture represented a 60:40 mixture of benzyldimethyldodecylammonium chloride and benzyldimethyltetradecylammonium chloride (C12BDMA-Cl and C14BDMA-Cl, respectively; Sigma-Aldrich). The three communities were maintained for four years at room temperature in an aerobic fed-batch 2-L Pyrex® reactor with a total liquid volume of 1.6L, a residence time of 14 days, and the previously described mineral medium plus carbon and energy source (i.e., 2200 mg/L of dextrin/peptone and 50 mg/L of BAC mixture). To obtain isolates, a mixed community suspension from each bioreactor (DP, DPB, and B) was first diluted and plated on agar medium containing dextrin/peptone, salt medium, and 1.5% agar as described previously (no neutralizer for BAC was used during isolation or transferring of isolates) (100). Taxonomic identification of the isolates was determined based on sequencing their 16S rRNA genes.

2.3.2 Construction of plasmids carrying efflux pump genes

Efflux pumps genes found on the integrative and conjugative elements (ICE) were cloned from *Achromobacter* B isolate and these genes were identical, at the nucleotide level, to those in *P. aeruginosa* DPB. The genes, including their native promoters, were individually amplified by PCR using primers containing enzyme restriction sites (EcoRI-BamHI for *sugE*-A: *sugE*-A-EcoRI: 5'-ATGCGAATTCCAGATAAAGCCAACCTTCC-3' and sugE-A-BamHI: 5'-ATGCGGATCCGACTACGCTACCAATGGAG-3' and *sugE*-B: *sugE*-B-EcoRI: 5'-ATGCGAATTCCATTATGGAAAGGGATGGCG-3' and

sugE-B-BamHI: 5'-ATGCGGATCCTTGCTCTTCATAATGGGTCTC-3' and KpnI-EcoRI for the ABC transporter operon: ABC-KpnI: 5'-ATGCGGTACCGTACTAGCGTCATAGTCACGG-3' and ABC-EcoRI: 5'-ATGCGAATTCCAACGTCATTAAAGAGTTCGC-3'). Digested amplicons were ligated into the multiple cloning site of pBBRMCS-4 (101). The resulting constructs were introduced into *P. aeruginosa*-PAO509, which was deficient for several RND-family efflux pump genes (*mexAB-oprM*, *mexCD-oprJ*, *mexEF-oprN*, *mexJK*, and *mexXY*) (102), by electroporation as described previously (103). Transformants were selected on LB agar medium supplemented with 100 μg/mL of ampicillin. Recombinant plasmids extracted from transformants were confirmed by PCR-amplification of inserts and enzyme restriction digestion.

2.3.3 Adaptive evolution experiments

A single colony of the *P. aeruginosa* DP and DPB isolates was used as the original inoculum for the adaptive evolution experiments. Six test tubes with a range of BAC concentration were initially inoculated with 1% aliquot of *P. aeruginosa* DP and DPB culture grown in a 10 ml Luria broth medium (LB) for 24 hours in triplicate at 35°C, using an orbital shaker at 225rpm (Figure S1B). Subsequently, 1% of the adapted population from the tube that showed growth (i.e., ≤ 80% growth reduction) at the tube with the highest concentration of BAC was transferred to new medium daily, followed by similar rounds of increasing BAC concentrations until no growth was observed in any tube [BAC(+) populations]. Minimum inhibitory concentration (MIC) was determined as the concentration in which ≥ 80% inhibition in cell growth was observed by OD 600 nm values compared to control (no BAC in the medium). A 1% aliquot of inoculum culture was used for each MIC measurement. Generation numbers of the cultures were measured every 24 hours by counting colony-forming units (CFUs). Inoculum from the tube that

showed growth in the highest BAC concentration was subsequently transferred to LB [BAC(-) populations], followed by rounds of growth in decreasing BAC concentrations. The BAC used for the experiment consisted of a 60:40 mixture of benzyldimethyldodecylammonium chloride and benzyldimethyltetradecylammonium chloride (C_{12}BDMA-Cl and C_{14}BDMA-Cl, respectively; Sigma-Aldrich).

2.3.4 Antimicrobial susceptibility test

The antimicrobial susceptibility test was performed with the microdilution procedure (104) in LB medium. All tested inocula were sampled at the end of the exponential phase and diluted in the medium to a final concentration of 5×10^5 CFU/mL before the test. Specifically, we characterized the growth curve for all the tested isolates, populations, and transformants with optical density values at 600 nm, and sampled at the end of exponential phase to measure CFU and optical density values at 625nm for 10-fold serially diluted aliquots (i.e., 10^{-1}, 10^{-2}, and 10^{-3}). Based on these data, the relationship between expected CFU and optical density values at 600 nm was determined, and this relationship was used to obtain aliquots of 5×10^5 CFU/mL based on the optical density values. Therefore, comparable inocula in terms of starting cell numbers were used in all cases. All tests were performed in triplicate on 48-well plates after growth of the inoculum (5×10^5 CFU/mL) for 24 hours at 35°C while mixed using an orbital shaker at 225 rpm. After incubation, the optical density values at 600nm were measured and the MIC was determined based on the control cultures with LB only medium for LB + antibiotics tests and LB+BAC medium for LB + BAC + antibiotics tests.

2.3.5 DNA extraction, RNA extraction and sequencing

Mixed community suspensions from bioreactors and isolates were taken for DNA extraction and processed as previously described (105). These DNA samples were

sequenced using the Illumina HiSeq 2000 sequencer at the Los Alamos National Laboratory Genomics Facility and reported previously (105); the ancestor *P. aeruginosa* DPB genome was closed using data from an in-house MinION instrument and the Illumina assembly. DNA of all *P. aeruginosa* populations from adaptive evolution experiments was extracted using the QIAamp DNA Blood Mini kit (Qiagen, Germany). DNA sequencing libraries were prepared using the Nextera XT DNA library prep and sequenced on an in-house Illumina MiSeq instrument (School of Biological Sciences, Georgia Institute of Technology) as described previously (106).

RNA of *P. aeruginosa* DPB populations from the adaptive evolution experiment was extracted as previously described (107) with slight modifications. In brief, culture suspensions (25 mL of LB culture, and 50 mL of LB+BAC culture) were collected during mid-log phase growth and centrifuged at 5000 x g for 10 minutes. BAC (1000 mg/L) was used in the growth medium for all DPB BAC(+) and BAC(-)_2 populations, and 100 mg of BAC/L was used for the ancestor, control, and BAC(-)_1 populations, which represented sub-inhibitory concentrations for the corresponding populations. The collected cell pellet was subsequently washed with 1X PBS, and re-suspended with 5 mL of lysis buffer (50 mM Tris-HCl, 40 mM EDTA ph 8, and 0.75 M sucrose) containing 15 mg/mL of lysozyme, and incubated at 37°C for 3 min. The lysates were then incubated with 0.4 mL of the lysis buffer, containing 1% SDS and 20 mg/mL of proteinase K, for 2 hours at 55°C in a rotating hybridization oven. RNA was extracted with acid phenol and chloroform and isolated by mirVANA RNA isolation kit (Ambion). RNA samples were treated with DNase using the TURBO DNA-free kit (Ambion, Austin, TX). Elimination of contaminating DNA was confirmed by PCR amplification of the *sugE*-A gene. Ribosomal RNA (rRNA) was depleted from total RNA using the Ribo-Zero rRNA Removal Kit (Illumina, San Diego, CA) and cDNA libraries were constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina, San Diego, CA) according to

manufacturer's instructions. cDNA libraries were quantified and sequenced as described above for *P. aeruginosa* population DNA samples.

2.3.6 Community metagenome and isolate genome sequence analysis

Raw Illumina reads were trimmed using a Q=15 PHRED quality score cutoff using SolexaQA(108) for further analysis. Community and isolate genome reads were assembled using the hybrid protocol previously described (109). Protein-coding genes encoded on assembled metagenomic or isolate genome contigs were annotated with the MetaGene pipeline (110). The phylogenetic affiliation of assembled contigs or unassembled metagenomic reads were determined based on best match searches against all bacterial and archaeal genome sequences available in GenBank database (as of June 2015, ftp://ftp.ncbi.nih.gov/), using Blastn (X = 150, q = -1, and F = F, remaining parameters at default settings) with a cutoff of a match with >95% identity and 50% query length coverage. The functional annotation of protein-coding genes was performed using a Blastp search against the SEED protein database (111), with a cutoff for a match of >40% amino acid sequence identity and >50% query length coverage.

We used Bowtie 2 (112) to map metagenomic reads from DPB bioreactor against *P. aeruginosa* DPB assembled contigs to estimate coverage, which was also taken as a proxy for relative abundance. To identify SNPs and genomic modifications, *P. aerginosa* DP genomic reads were matched against *P. aeruginosa* DPB assembled contigs, and vice versa using breseq (consensus mode and Q=15 PHRED quality score cutoff were used) (113, 114). The same approach was also applied in comparisons of *P. aeruginosa* DP and DPB control, BAC(+), and BAC(-) populations against *P. aeruginosa* DP and DPB assembled contigs (ancestors) in the adaptive evolution experiments. Identified mutations in the evolved control population against its ancestor were considered to be the result of stochastic processes or selection by the growth conditions (e.g., "bottle effect"), and were

removed from further analysis. Read recruitment plots were obtained as described previously (115) based on the results of a Blastn search of reads against the GI-2 sequence of *P. aeruginosa* DPB and a minimum cut-off for a match of 70% of identity and 60 bp alignment length.

2.3.7 MinION sequencing and read analysis

Sequencing libraries for the MinION sequencer were prepared using the SQK - RAD003 Rapid Sequencing Kit, R9 version lot SR03.10.002 (Oxford Nanopore Technologies, Oxford, United Kingdom). Approximately, 421 ng of *P. aeruginosa* strain DPB DNA was tagmented by incubation with 2.5 µL Fragmentation Mix (FRM) for 1 min at 30 °C followed by 1 min at 80 °C. Tagmented DNA was then adapter - ligated by incubation with 1.0 µL Rapid Adapter mix (RAD) for 5 min at room temperature and stored on ice. A FLO-MIN 107 (R9.5) flow cell was primed as directed by the manufacturer and then loaded with a mixture of 11 µL tagmented library, 30.5 µL Running Buffer FM (RBF), 26.5 µL Library Loading Beads (LLB) and 7 µL Nuclease-free water. *P. aeruginosa* DPB was sequenced using the SQK-RAD003 MinKNOW protocol script with local basecalling and the sequencing was stopped after 12 hours.

After base-calling of the 1D MinION reads, the reads were extracted in fastq format using Poretools version 0.6.0 (116). All resulting reads were trimmed using NanoFilt (https://github.com/wdecoster/nanofilt) with a PHRED score cutoff of 10 and a minimum fragment length of 1 Kbp, and the adapter sequences were removed using Porechop (https://github.com/rrwick/Porechop). Then, the hybrid read set, i.e., Illumina and MinION trimmed reads, were assembled using Unicycler v0.4.2 with the conservative mode (117). Briefly, Unicycler uses a SPAdes assembly of Illumina reads and then scaffolds the assembly graph using MinION reads. Finally, Unicycler uses Illumina reads to polish the final assembly using Pilon (118). To identify genomic islands

specific to the *P. aeruginosa* DPB closed genome from the hybrid Illumina+MinION assembly, *P. aerginosa* DP genomic reads (Illumina) were searched against the *P. aeruginosa* DPB genome using breseq (consensus mode and Q=15 PHRED quality score cutoff were used) (113, 114) and the genomic regions not covered by reads were considered strain DPB-specific.

2.3.8 Transcriptome sequence analysis

All transcriptomic reads of *P. aeruginosa* DPB ancestor, control population, all BAC(+), BAC(-)_1 and BAC(-)_2 populations were trimmed using SolexaQA (108) with a PHRED score cutoff of 20 and a minimum fragment length of 50 bp, and 3'-end adapter contaminants were removed using Scythe v0.993 (https://github.com/vsbuffalo/scythe). These trimmed reads were filtered using SortMeRNA v2.0(119) to remove remaining rRNA sequences with all databases in the program. Most of the resulting datasets had >95% of their total reads to be non-rRNA except three libraries (Table A 7). One library (i.e., *P. aeruginosa* DPB BAC(+)_2L_3) in particular showed 9.42% of reads to be non-rRNA which resulted in too few non-rRNA sequences relative to the remaining libraries (e.g., >2 fold difference in number of non-rRNA reads), and thus was removed from further analysis in order to not bias the comparisons and results (120). Non-rRNA reads were mapped to *P. aeruginosa* DPB assembled contigs using Bowtie 2 (112) and read count tables against predicted genes were generated by featureCounts v1.4.6-p3 (121). The output read count tables were used as input for DESeq2 (122) to obtain the lists of differentially expressed genes in pairwise comparisons. The pairwise comparisons included both the type of populations (e.g., ancestor vs. BAC(+) populations) and culture conditions (i.e., BAC-free vs. BAC-supplemented). All genes that were differentially expressed from all pairwise comparisons with cutoff of padj < 0.01 and number of read counts > 1 were included for

further analysis. Differentially expressed genes from the comparison between control population and ancestor that were also observed in the comparison of BAC(+) or BAC(-) populations against the ancestor were discarded from further analysis, similar to the mutations described above.

2.3.9 Data release

The bioreactor metagenome sequences used in this study were deposited in GenBank under the accession numbers: SRR643889 (DP), SRR643891 (DPB), and SRR643892 (B). The genome sequences of isolates, all genomic and transcriptomic sequences of *P. aeurginosa* DP/DPB BAC(+) and BAC(-) populations, and MinION sequencing data can be found under the BioProject PRJNA184698.

2.4 Results and Discussion

2.4.1 Microbial community exposure to BAC selected for members with increased resistance to antibiotics.

To examine whether BAC exposure promoted (i.e., increased) antibiotic resistance, pairs of isolates, one originating from the DPB or B (i.e., BAC-fed) bioreactor and the other one representing its counterpart (same species) from the DP (i.e, no-BAC-fed) bioreactor, were characterized for their resistance to representatives of seven classes of antibiotics. The isolates represented four distinct species that were relatively abundant in the bioreactors and highly represented among the total isolates obtained (Figure A 1). *P. aeruginosa* strain DPB showed higher minimal inhibitory concentration **(**MIC) values for BAC and several, but not all, antibiotics compared to *P. aeruginosa* strain DP (Table 2-1). The genomes of these two *P. aeruginosa* isolates differed in their shared genes by

five single nucleotide polymorphisms (SNPs), one small (9 bp) insertion in their shared genes and two genomic islands (GIs) (Figure 2-1). This number of mutations was comparable to the number of mutations predicted (n= ~10) to be fixed under neutral evolution based on the spontaneous mutation rate for a bacterial genome (5.4×10^{-10} per base per generation) (123), the genome size of the *P. aeruginosa* isolates (~6.4 Mb), and the estimated number of generations since the establishment of the DPB and DP communities after three years (~3,000 generations). Hence, it is highly likely that these isolates represented descendants of the same ancestor in the original river sediment inoculum. *P. aeruginosa* can grow in soil, marshes, coastal marine habitats, plants, and animal tissues (124) and it is commonly isolated from sediments (125). Therefore, it was not surprising to recover *P. aeruginosa* from our sediment inoculum.

Similar antibiotic resistance results to *P. aeruginosa* isolates were observed for *Achromobacter sp.* strain DPB/B and *Achromobacter denitrificans* strain DP, which represented more divergent genotypes than the *P. aeruginosa* isolates (genome-aggregate average nucleotide identity or ANI between the *Achromobacter sp.* DPB/B and *A. denitrificans* DP isolates was 87.5%, Table 2-1). In contrast, *Citrobacter freundi* strain DPB and *Klebsiella michiganensis* strain DPB did not exhibit significant changes in MIC values for antibiotics compared to their unexposed counterparts (Table 2-1). However, the latter two isolates showed higher MIC values for BAC and several antibiotics compared to *C. freundii* and *K. michiganensis* isolates from culture collections (type strains) or the human gastrointestinal tract, indicating that *C. freundii* DP and *K. michiganensis* DP might have been intrinsically resistant to several of the antibiotics tested here before BAC exposure. Overall, although BAC exposure did not result in

higher antibiotic resistance in all isolates tested, at least *P. aerugionsa* showed reduced susceptibility to BAC by 4 fold and three antibiotics from the seven tested including ciprofloxacin, chloramphenicol, and rifampicin by 2 fold, suggesting a resistance link between BAC and antibiotics.

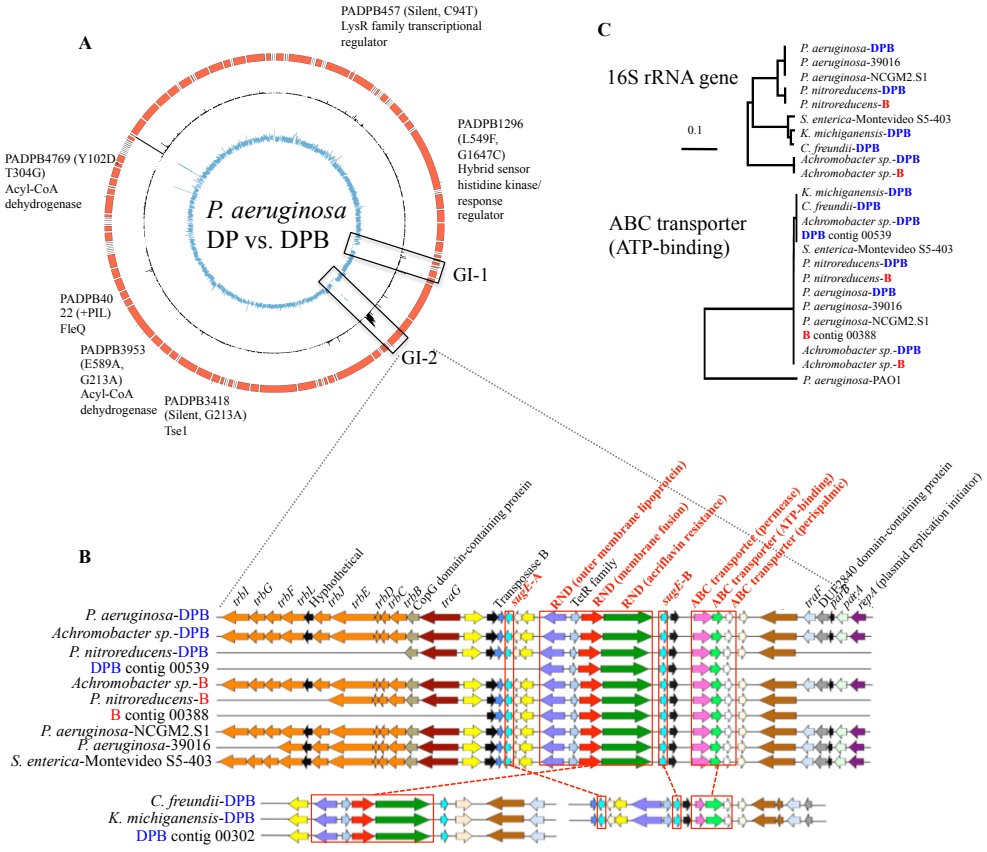


Figure 2-1. Horizontal transfer of antibiotic efflux pump genes. (A) Comparison of the draft genomes of *P. aeruginosa* DPB vs. *P. aeruginosa* DP. Circles represent: contigs (red) of *P. aeruginosa* DPB and coverage of contigs by metagenomic reads of DPB reactor (black), and the *P. aeruginosa*-DP strain (blue). Genes with mutations found between the two genomes are denoted on the graph (ticks on the *P. aeruginosa* DPB draft genome). (B) Graphic representation of the organization of genes encoded on the ICE recovered from isolate genomes and metagenomes. (C) Phylogenetic tree of 16S rRNA gene sequences and ATP-binding gene (of the ABC transporter) sequences based on the maximum likelihood composite model and the neighbor-joining method. Note that

phylogenetically diverse isolates shared almost identical (>99% identity) ATP- binding gene sequences, suggesting horizontal transfer of the ATP-binding gene. Similar results were also observed for *sugE*-A and *sugE*-B genes (data not shown). The reference sequences used were *P. aeruginosa*-39016 (accession no. NZ_CM001020.1), *P. aeruginosa*-NCGM2.S1 (NC_017549.1), *P. aeruginosa*-PAO1 (NC_002516.2), and *Salmonella enterica*-Montevideo-S5-403 (AFCS01000001.1).

Table 2-1. MIC of BAC and antibiotics for isolates of same species. A range of antibiotic and BAC concentrations (0.4 to 3200 mg/L) were tested and the table shows the determined MIC values, 16S rRNA gene sequence identity and average nucleotide identity (ANI) against the control isolate of the same or closely related species (denoted by "-"). C. freundii Human strain was 4_7_47_CFAA, accession no. ADLG00000000.1, of the Human Microbiome Project (http://www.hmpdacc.org/) and K. michiganensis type strain was KCTC 1686, accession no. CP003218.1, of Korean Collection for Type Cultures. a TET, Tetracycline; CIP, Ciprofloxacin; CHL, Chloramphenicol; POL, Polymyxin B; KAN, Kanamycin; RIF, Rifampicin; AMP, Ampicillin

Species	16S rRNA gene identity (%)	Average nucleotide identity (ANI, %)	MIC, mg/L[a]							
			BAC	TET	CIP	CHL	POL	KAN	RIF	AMP
Achromobacter denitrificans DP	-	-	15.6	< 3.1	< 0.8	3.1	< 0.4	100	12.5	200
Achromobacter sp. DPB	97.8	87.5	31.3	25	1.6	25	< 0.4	25	3.1	< 6.3
Achromobacter sp. B	97.7	87.5	125	200	12.5	100	3.1	100	12.5	< 6.3
Citrobacter freundii Human	-	-	< 7.8	1.6	0.4	12.5	0.2	6.3	25	800
Citrobacter freundii DP	98.5	N/A	62.5	1.6	1.6	12.5	0.8	6.3	25	12.5
Citrobacter freundii DPB	99.5	94.2	62.5	1.6	0.8	12.5	0.4	6.3	25	25
Klebsiella michiganensis type	-	-	15.6	1.6	< 0.1	6.3	1.6	200	12.5	200

Klebsiella michiganensis DPB	99.9	99.4	62.5	1.6	1.6	12.5	1.6	25	25	100
Pseudomonas aeruginosa DP	-	-	50	12.5	0.1	50	0.2	100	12.5	3200
Pseudomonas aeruginosa DPB	100	100	200	6.25	0.2	100	0.2	100	25	3200

2.4.2 Genetic elements encoding genes for resistance to both BAC and antibiotics were responsible for the link.

To investigate the molecular mechanisms responsible for the resistance link, we conducted whole isolate and bioreactor metagenome DNA sequencing. Alignment of the *P. aeruginosa* DP genomic reads against the *P. aeruginosa* DPB assembled contigs identified, in addition to the five SNPs and the small insertion mentioned above, two GIs that encoded several integrases and transposases (GI-1) together with several (predicted) resistance genes (GI-2) (Figure 2-1A). GI-2 harbored four predicted efflux pump systems: two small multidrug resistant (SMR) family systems, (*sugE*-A and *sugE*-B), an ATP-binding cassette (ABC) family, and a resistance nodulation division (RND) family member (Figure 2-1B). Furthermore, GI-2 encoded all necessary genes for conjugation such as conjugative transfer (*trbBCDEJLFGI*), plasmid partitioning genes (*parAB*), plasmid replication initiator (*repA*) in a single assembled Illumina contig. Recruitment of Illumina reads from the DPB metagenome revealed that the coverage of GI-2 (Avg. 17.7 reads/base pair or 17.7X) was significantly higher ($P < 0.001$, Student's t-test) than the rest of the *P. aeruginosa*-DPB genome (Avg. 2.0X), indicating that GI-2 was present in additional members of the DPB community or in multi-copy (Figure 2-1A). Consistent with this interpretation, genome sequence analysis of *Achromobacter sp.*, *P. nitroreducens*, *C. freundii*, *K. michiganensis* and *P. aeruginosa* isolates from DPB or B communities revealed an almost identical (>99% nucleotide identity) genetic element despite substantial evolutionary divergence of the organisms (Figure 2-1B and Figure 2-1C). Since GI-2 was the only genomic difference between *P. aeruginosa* DP and DPB strains that was bioinformatically-predicted to encode enzymes potentially relevant for

BAC and antibiotic resistance, we hypothesized that GI-2 might have been responsible for the higher antibiotic resistance of strain *P. aeruginosa* DPB relative to strain DP and thus, we studied this genomic island in more detail.

Even though GI-2 was predicted to encode several of the genes commonly found in conjugative plasmids, long-read Oxford Nanopore sequencing showed that the conjugative element was integrated in the genome of *P. aeruginosa* DPB. The hybrid assembly of the MinION long reads and Illumina contigs produced a closed genome (single contig), with even coverage across the genome, including the junctions/ends of the Illumina contig encoding the conjugative transfer genes. Further bioinformatics analysis and gene annotation, e.g., presence of recombinase genes (often referred to as integrases) and conjugation system genes (Table A 1) (126, 127), suggested that the GI-2 is most likely an ICE. ICEs share gene content with conjugative plasmids (e.g., genes for conjugation), and are self-transmissible genetic elements that can be integrated in the chromosome or propagate as independently replicating molecules (See Supporting results and discussion for further details) (126, 127). Alignment of genomic reads of other isolates, i.e., *Achromobacter sp.* DPB/B, *C. freundii* DPB and *K. michiganensis* DPB, against *P. aeruginosa* DPB GI-2 sequence suggested that not all of the isolates shared the whole GI-2 or all integrases, except for *Achromobacter sp.* DPB (**Figure A 2**). These results revealed that GI-2 has likely been mobilized and/or integrated into diverse organisms during bioreactor incubation time (3 years), and various versions of ICEs encoding the same resistance genes were present within the community, which were apparently strongly selected by the BAC exposure conditions.

To test the hypothesis that GI-2-encoding genes helped organisms to cope with BAC toxicity and were responsible for the higher antibiotic resistance of *P. aeruginosa* strain DPB relative to strain DP, the *sugE*-A, *sugE*-B, and ABC transporter system genes were cloned into a broad-host range vector (pBBRMCS-4) and transformed into *P. aeruginosa*-PAO509. PAO509 lacks several efflux pump systems present in the wild-type *P. aeruginosa* and our isolates from the bioreactors and hence, is a more appropriate strain background to test for antibiotic resistance phenotypes. Plasmid pBBRsugE-A conferred increased resistance to BAC, with a ~2 fold higher MIC, but not antibiotics, while plasmid pBBRABC conferred resistance to rifampicin (~2 fold higher MIC) but BAC tolerance was similar to the control vector alone (pBBRMCS-4) (Table A 2). It should be noted that even a two fold higher MIC can be clinically significant, and efflux pumps genes such as *sugE*-A and ABC do not often provide for much greater changes in MIC values (128-131). The BAC-resistance phenotype of the transformant carrying *sugE*-A (PAO509/pBBRsugE-A) was also in agreement with a previous report that *sugE* confers resistance to BAC (132). Rifampicin is known to act on the RNA polymerase β subunit (RpoB) to interfere with transcription and rifampicin-resistant bacteria frequently emerge due to single point mutations in *rpoB* (133). To exclude the possibility that the rifampicin-resistant phenotype was due to spontaneous mutation (as opposed to the ABC transporter system cloned), *rpoB* gene sequences of Rif clusters I, II, and III were PCR-amplified from four independent colonies of PAO509 and PAO509/pBBRABC using previously determined primers PAO1rpoB1 (134) and PAOrpoB3 (135) and subsequently sequenced using Sanger chemistry. No mutation was identified between PAO509 and PAO509/pBBRABC, suggesting that the rifampicin resistance phenotype was likely not

due to spontaneous mutation but to the ABC transporter system. No difference in tetracycline susceptibility was observed in the three recombinants compared to the control. These results suggested that two distinct gene cassettes encoded by GI-2 conferred resistance to antibiotic and BAC respectively, but no one gene conferred resistance to both BAC and antibiotics. Therefore the resistance linkages appeared to be due to the co-occurrence of these two gene cassettes on the same piece of mobile DNA. It still remains unclear at this point if the co-localization of the genes occurred before or after exposure to BAC.

2.4.3 BAC-adaptation selected for P. aeruginosa with increased resistance to polymyxin B.

To identify alternative genetic mechanisms that may be responsible for the resistance linkage observed but were not elucidated by our genetic manipulations, and assess the effects of shorter BAC exposure compared to that experienced by the microbial communities in the DPB and B bioreactors (~3 years), the two *P. aeruginosa* isolates DP and DPB were subjected to increasingly higher BAC concentrations every 24 hours, in triplicate batch cultures, until no growth was observed. The last (highest) concentration for which growth was observed was recorded for each replicate population [BAC (+)]. Subsequently, the populations were transferred and maintained in BAC-free medium [BAC(-); **Figure 2-2**A, **Figure 2-2**B, and Figure A 1B]. During this evolution experiment, the two genotypes existed on different adaptive landscapes because strain DPB possessed the ICE-encoding resistance genes and therefore, was inherently more resistant to BAC (Table 2-1). The other pairs of isolates recovered that showed cross-resistance such as *Achromobacter* were not used in the adaptive evolution experiment

because these experiments were laborious and the pair of genomes recovered were divergence from each other (i.e. having different genomic backgrounds), which could complicate results and interpretations.

All DP and DPB populations showed increased resistance to BAC after growth with BAC for about 240 and 180 generations, respectively, compared to their ancestor (7 to 25 fold higher MIC) or the controls (*P. aeruginosa* DP or DPB population evolving in parallel, under the same laboratory conditions but with no BAC in the growth medium). No BAC biodegradation by any of the populations was observed while the highest BAC tolerance observed (by DPB population) was 1,600 mg/L BAC. After transfer to BAC-free medium, the three DP BAC(-) populations lost their BAC tolerance within 150 generations, and their BAC MIC was similar to that of the ancestor (50 mg/L BAC), indicating that adaptation to BAC might have been transient, at the level of cell physiology and gene regulation, or due to the rise of compensatory mutations that restored sensitivity to BAC under growth in BAC-free medium. In contrast, two of the DPB BAC(-) populations maintained high BAC tolerance with MIC for BAC similar to those of the BAC(+) populations and about 7 fold higher than that of their ancestor while the third DPB BAC(-) population lost the acquired BAC tolerance. These results indicated that the two DPB BAC(+) populations might have acquired genetic modifications upon BAC exposure that were maintained in their corresponding DPB BAC(-) populations. Since all DP and DPB BAC (+) populations, and two of the DPB BAC(-) populations acquired increased tolerance to BAC, these populations were examined further to assess if the increased BAC tolerance was accompanied by increased resistance to antibiotics and the underlying mechanism(s).

Interestingly, we observed that all DP BAC(+) populations showed higher MIC (2 to 8 fold) to the membrane-active antibiotic polymyxin B compared to the ancestor or the control (Table 2-2). Of the remaining antibiotics tested, DP BAC(+) populations did not show higher MIC than their ancestor (**Table A 3**A). To identify the molecular mechanisms for cross resistance between BAC and polymyxin B, whole genome sequencing was conducted and identified 29 SNPs and 17 DIP (deletions, insertions and other polymorphisms) in *P. aeruginosa* DP BAC(+) and BAC(-) populations relative to the control population, and 22 SNPs and 16 DIP in *P. aeruginosa* DPB BAC(+) and BAC(-) populations, respectively (Table A 4A, Table A 4B, Table A 4C and Table A 4D). Remarkably, all DP BAC(+) and DPB BAC(+) populations had fixed mutations compared to their ancestral populations in only a single gene found in common, *pmrB*, albeit at different positions of the gene. *pmrAB* encodes a two-component regulatory system, and mutations in *pmrB* are known to confer polymyxin resistance in both clinical isolates and a laboratory strain of *P. aeruginosa* via constitutive activation of the *pmrA* regulon (136). Expression of *pmrA* leads to the expression of *arnBCADTEF*, which is responsible for the addition of 4-amino-L-arabinose (L-Ara4N) to the phosphate groups of lipid A, resulting in reduction of net negative charge on the outer membrane (Figure 2-3B) (137, 138). Because both polymyxin B and BAC are cationic membrane-disrupting agents (139, 140), the reduction in net negative charge is expected to increase BAC and polymyxin B tolerance by reducing their adsorption to the outer membrane. Therefore, our results revealed that mutations to the *pmrB* gene in the DP BAC(+) populations were selected by BAC exposure and led to increased resistance to at least one antibiotic, polymyxin B.

It should be noted, that although all the *P. aeruginosa* populations had fixed mutations in *pmrB*, only the DP BAC(+) populations showed increased resistance to polymyxin B compared to the DP BAC(-), DPB BAC(+) and DPB BAC(-) populations under our experimental conditions (Figure 2-3A, Table 2-2, and **Table A 3**B). Read alignment of the datasets of DP BAC(-) populations against assembled contigs of the ancestors revealed that the populations had different mutations in *pmrB* and *pmrA* compared to those in DP BAC(+) populations (**Table A 5**). RNA-seq results suggested that the mutations to *pmrB* in DPB BAC(+) and DPB BAC(-) populations did not lead to the overexpression of *arnBCADTEF* relative to the ancestor and control (See Supporting results and discussion for further details). These findings presumably explained the lack of increased polymyxin B resistance in the DP BAC(-), DPB BAC(+) and DPB BAC(-) populations because the location and combination of mutations in *pmrB* are known to affect the level of expression of *arnB* (137), as confirmed by our RNA-seq data, and thus, polymyxin B resistance. Further, repeated passages without polymyxin in the growth media have been shown to result in loss of resistance in some cases, suggesting that the *pmrAB* locus is not the only determinant of the resistance phenotype (137) (See Supporting results and discussion for further details). The lack of increased polymyxin resistance and the different mutations acquired in the *pmrB* gene in DPB populations are also not surprising given that these populations existed on different adaptive landscapes at the beginning of the experiment than DP populations. Altogether, BAC exposure selected for the mutations in *pmrB* that conferred increased BAC-tolerance in all *P. aeruginosa* populations, but only specific mutations conferred increased polymyxin B resistance in the DP (but not the DPB) genetic background.

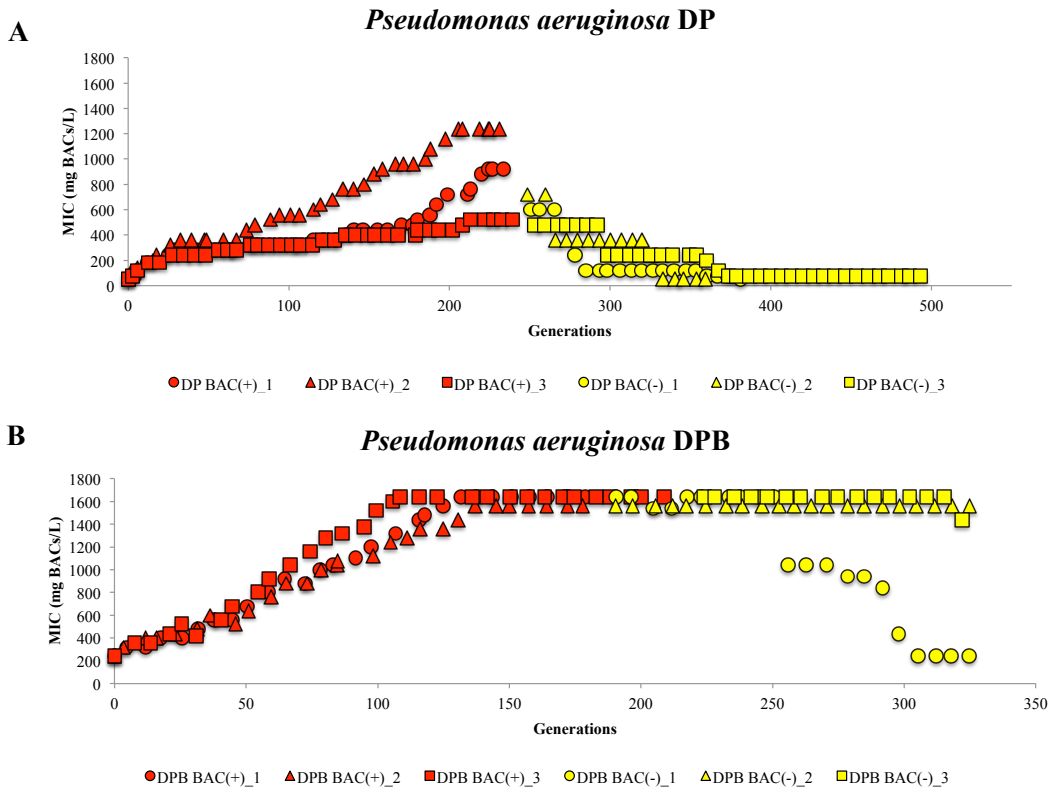


Figure 2-2. Changes in minimal inhibitory concentration (MIC) of BAC during an adaptive evolution experiment with *P. aeruginosa* DP (A) and *P. aeruginosa* DPB (B). The adaptive evolution experiment consisted of two phases; BAC-exposure, where BAC concentration was increased after each round until reaching the maximum concentration in which growth was observed (red dots; each dot represents a passage), and BAC-free (yellow dots). The MIC for BAC (y-axis) was measured during each round of growth, for about 320 to 500 generations. The X-axis shows the number of generations since the start of the experiment.

Table 2-2. MIC values for polymyxin B of all populations of *P. aeruginosa* DP from the adaptive evolution experiment

P. aeruginosa *DP* populations (LB cultured)	MIC, mg/L
	Polymyxin B
BAC(+)_1	1.6
BAC(+)_2	0.8
BAC(+)_3	0.8
Control_1	0.4
BAC(-)_1	0.2
BAC(-)_2	0.4
BAC(-)_3	0.4
Ancestor	0.2
Control_2	0.4

Table 2-3. MIC values for tetracycline and ciprofloxacin tested with no-BAC and added-BAC growth condition for *P. aeruginosa* DPB evolved populations

P. aeruginosa DPB populations	MIC, mg/L			
	Tetracycline	Ciprofloxacin	Tetracycline +BAC(100mg/L)	Ciprofloxacin +BAC(100mg/L)
BAC(+)_1	12.5	0.2	25	3.2
BAC(+)_2	3.125	0.2	12.5	1.6
BAC(+)_3	12.5	0.2	25	0.8
Ancestor	6.25	0.2	25	1.6
Control_1	12.5	0.4	25	1.6

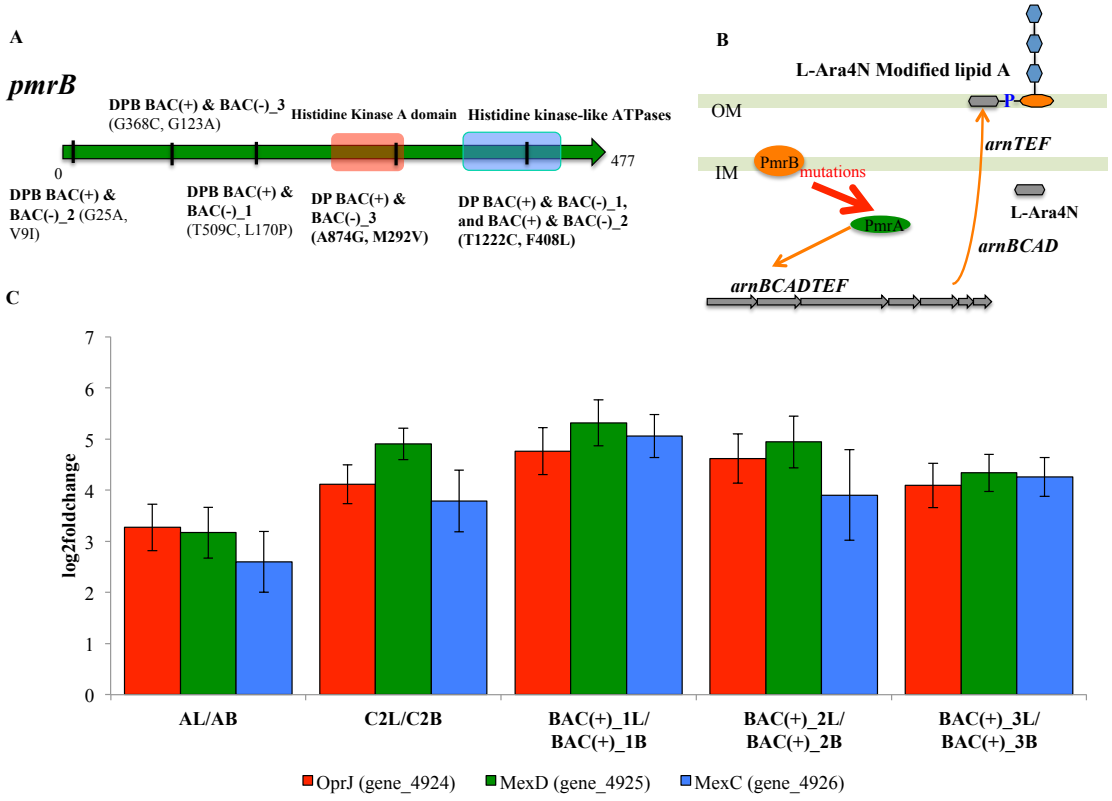


Figure 2-3. Mutational and transcriptional evidence for the resistance links between BAC and antibiotics. (A) Mutations in *pmrB* genes at various positions were noted for each population (denoted on the graph). (B) Schematic of the known polymyxinB resistance mechanisms caused by mutations in *pmrB* gene(136, 138, 141). (C) Overexpression of *mexCD-oprJ* operon when BAC is supplemented in the growth media. AL/AB: *P.aeruginosa* DPB ancestor in LB growth media versus LB+BAC media; C2L/C2B: *P. aeruginosa* DPB_Control_2 in LB vs. LB+BAC; BAC(+)_1L/BAC(+)_1B: *P. aeruginosa* DPB BAC(+)_1 in LB vs. LB+BAC; BAC(+)_2L/BAC(+)_2B: *P. aeruginosa* DPB BAC(+)_2 in LB vs. LB+BAC; BAC(+)_3L/BAC(+)_3B: *P. aeruginosa* DPB BAC(+)_3 in LB vs. LB+BAC.

2.4.4 Physiological adaptation to BAC exposure also led to increased antibiotic resistance.

During this evolution experiment, another important resistance link was revealed. All *P. aeruginosa* DPB ancestor, BAC(+), and control populations showed increase

resistance to tetracycline and ciprofloxacin (2 to 16 fold) when BAC was in the growth media at sub-inhibitory concentrations during the MIC test (Table 2-3 and **Table A 3**C). All populations tested here reached at least ~10^7 cells/mL after 24 hours growth at 35°C in the presence of sub-inhibitory concentrations of BAC, and these populations were used as controls for determining MIC values for antibiotics. Therefore, the sub-inhibitory concentration of BAC used did not inhibit growth, and the MIC values obtained were presumably not due to experimental artifacts such as absorbance contributed by dead cell debris. Thus, it appeared that co-presence of (sub-inhibitory) BAC in the growth medium of *P. aeruginosa* DPB induced physiological changes to the exposed cells that also conferred antibiotic resistance regardless of the level of BAC tolerance of the population. RNA-seq analysis revealed, for instance, over-expression (6 to 40 fold) of the *mexCD-oprJ* multidrug efflux pump (Figure 2-3C) under BAC-supplemented compared to BAC-free growth conditions. These findings were also consistent with previous results that *mexCD-oprJ* contributes to resistance to several classes of antibiotics such as fluoroquinolones (e.g., ciprofloxacin, and norfloxacin) and tetracycline (142, 143). Furthermore, previous studies have shown that clinically relevant disinfectants such as BAC but not antibiotics (e.g., norfloxacin, tetracycline, or chloramphenicol) can induce the expression of this family of efflux pumps even though antibiotics such as tetracycline and norfloxacin (in the same class as ciprofloxacin) are substrates for the pump (144, 145), in agreement with the results reported here.

2.5 CONCLUSIONS AND OUTLOOK

Collectively, our results revealed that BAC exposure can induce antibiotic resistance via multiple genetic mechanisms, including co-occurrence of BAC-tolerance

and antibiotic resistance genes on the same mobile DNA molecule, mutations in the *pmrB* gene, and induction of efflux pump expression. These results are important because the cross-link between BAC exposure and antibiotic resistance was observed in at least three antibiotics of the seven tested (Table 2-1), the underlying mechanisms were elucidated (Figure 2-1 and Figure 2-3) so the cross-link can not be attributed to spurious findings, and even a two fold higher MIC as revealed for polymyxin B and rifampicin can be clinically significant based on the pharmacokinetics and pharmacodynamics of these two antibiotics (129, 130). Further, the concentrations of BAC that the adapted *P. aeruginosa* DP BAC(+) and DPB BAC(+) populations were able to withstand during the adaptive evolution experiments (up to 1,600 mg/L) were comparable to, or even higher, than those used in practice as disinfectant (typically between 400-500 mg/L, and almost always below 1,000 mg/L) (146). Therefore, our results suggested that accumulation of BAC in any non-target environments (e.g., freshwater or sediments habitats, or the waste stream of hospitals or food processing facilities) should be prevented in order to limit the spreading of antibiotic resistance determinants. New biotechnologies that employ the recently reported BAC-degrading organisms or their enzymes could be used for the latter purpose (98, 106). Our results also helped to explain why there is a debate in the literature on whether or not linkage exists between exposure to disinfectants and increased antibiotic resistance. In particular, multiple adaptive possibilities (e.g., mutations in *pmrB*) in *P. aeruginosa* isolates were observed in response to BAC exposure, which is perhaps not unexpected given the non-specific nature of the disinfectant, and some -but not all- of these mutations also conferred benefits under other conditions, specifically polymyxin exposure. Similarly, not all bacterial species exposed

to BAC acquired increased resistance to antibiotics but at least several did (e.g., *P. aeruginosa*). It is also important to note that the list of organisms or growth conditions studied here were, by no means, exhaustive and additional mechanisms, which may cause even higher antibiotic resistance levels than what was noted here, likely exist in natural or clinical/engineered environments. Therefore, more attention should be given to the usage of disinfectant and the better understanding of their fate, especially in non-target environments. Finally, our experimental design and approach used in the adaptive evolution of *P. aeruginosa* can be employed to assess whether or not other important pathogens also show cross-resistance for BAC and antibiotics, and identify the underlying molecular mechanisms.

2.6 ACKNOWLEDGMENTS

This work was supported by US National Science Foundation award CBET 0967130 to SGP and KTK and award 1241046 to KTK.

CHAPTER 3 GENOMIC AND TRANSCRIPTOMIC INSIGHTS INTO HOW BACTERIA WITHSTAND HIGH CONCENTRATIONS OF BENZALKONIUM CHLORIDE BIOCIDES

This chapter was published in:

Minjae Kim, Janet K. Hatt, Michael R. Weigand, Raj Krishnan, Spyros G. Pavlostathis, and Konstantinos T. Konstantinidis. *Appl. Environ. Microbiol.* 84.12 (2018): e00197-18.

3.1 Abstract

Benzalkonium chlorides (BAC) are commonly used biocides in broad-spectrum disinfectant solutions. How microorganisms cope with BAC exposure remains poorly understood, despite its importance for disinfection and disinfectant-induced antibiotic resistance. To provide insights into these issues, we exposed two isolates of an opportunistic pathogen, *Pseudomonas aeruginosa*, to increasing concentrations of BAC. One isolate was pre-adapted to BAC as it originated from a bioreactor fed with sub-inhibitory concentrations of BAC for 3 years, while the other originated from a bioreactor that received no BAC. Replicated populations of both isolates were able to survive high concentrations of BAC, up to 1200 and 1600mg/L for the non- and pre-adapted ones, respectively, exceeding typical application doses. RNA-seq analysis revealed up-regulation of efflux pump genes and decreased expression of porins related to BAC transport as well as reduced growth rate. Increased expression of spermidine (a polycation) synthase genes and mutations in the *pmrB* (polymyxin resistance) gene, which cause a reduction in membrane negative charge, suggested that a major adaptation

to exposure to the cationic surfactant BAC was to actively stabilize cell surface charge. Collectively, these results revealed that *P. aeruginosa* adapts to BAC exposure by a combination of mechanisms, and provided genetic markers to monitor BAC-resistant organisms that may have applications in the practice of disinfection.

3.2 Introduction

Benzalkonium chlorides (BAC) are among the most commonly used members of the quaternary ammonium compound (QAC) as biocides. BAC act as cationic surfactants due to their amphiphilic property, i.e., positively charged nitrogen atom in the center of a molecule with C_8 to C_{18} n-alkyl chain. BAC have broad-spectrum biocidal activity against microbial, algal, fungal, and viral organisms, based on the following modes of action (23): (i) adsorption and penetration of the BAC into the cell wall; (ii) interaction with the cytoplasmic membrane (lipid or protein) followed by membrane disorganization; (iii) leakage of intracellular light molecular weight material; (iv) degradation of proteins and nucleic acids; and (v) cell wall lysis by activity of autolytic enzymes (147, 148). BAC remain stable under various conditions in acidic, alkaline, and oxidative environments (149). Accordingly, BAC are widely employed in disinfectant solutions/formulations for food processing lines, such as poultry facilities and dairy/agricultural settings, health care facilities and domestic households, and are popular ingredients for over-the-counter products such as cosmetics, hand sanitizers, and pharmaceuticals (24). Their extensive use and stability, especially under sediment-sorbed conditions (but not under non-sorbed, aerobic conditions where BAC can be quickly biodegraded; see also below), have caused BAC concentrations to often be higher than those of other conventional organic pollutants (e.g., polyaromatic hydrocarbons) or other

QAC representatives in non-target environments (150, 151). BAC have been detected across diverse environments such as river (at concentrations of ~6 μg/kg) and estuarine sediments (1.5 μg/kg), surface water (1.9 μg/L), wastewater influent (170 μg/L) and hospital wastewater, where they can reach concentrations up to ~6.03 mg/L (150-153).

Although BAC disinfectant solutions applied in hospital settings have a pH that is neutral to alkaline, are noncorrosive, non-staining, and are considered safe to apply on all washable surfaces (154, 155), at least 40 outbreaks have been attributed to infection by disinfectant-resistant pathogens contaminating antiseptic products such as *Pseudomonas aeruginosa* (25-33, 35, 156). It has been hypothesized that the contaminating pathogens may be associated with the water used in pharmaceutical processing facilities, or the storage of BAC with cotton or gauze (35, 157). Thus, a better understanding of how microorganisms become resistant to BAC may help to better control such outbreak incidents. However, the underlying mechanisms of BAC-resistance and their relative importance remain poorly elucidated (157) (see below for details). Furthermore, there is rising concern that widely-used disinfectants such as BAC can promote antibiotic resistance (4, 17-21). For instance, our previous study revealed that exposure to BAC can promote resistance to various antibiotics, including polymyxin B and identified the molecular mechanisms for the resistance linkage between BAC and antibiotics (M. Kim, M. R. Weigand, S. Oh, J. K. Hatt, R. Krishnan, U. Tezel, S. G. Pavlostathis, and K. T. Konstantinidis, unpublished data). Antibiotic resistant infections cause millions of casualties annually, emphasizing the need to better understand BAC resistance mechanisms that can also contribute to antibiotic resistance (9).

Several studies have reported BAC resistance in phylogenetically diverse microorganisms, most often related to the activation of efflux pump genes from four families, i.e., the major facilitator superfamily (MFS) (36, 37), the multidrug and toxic compound extrusion family (MATE) (158, 159), the resistance nodulation cell division family (RND) (160), and the small multidrug resistance family (SMR) (161, 162). In addition to efflux pump up-regulation, dealkylation of BAC to benzyl dimethylamine (BDMA) and a long alkyl chain represents another important process for the development of BAC resistance (biodegradation) because BDMA is about 500 times less toxic than BAC (10, 97, 98). Furthermore, several studies have suggested that BAC resistant bacteria show cell envelope modifications such as reduced negative charge of the cell surface, increased cell surface hydrophobicity, reduced cell surface roughness, and increased saturated fatty acids components on the cell surface, probably resulting in the decrease of cell membrane permeability and fluidity (40, 163, 164). Because the cell wall in Gram-positive bacteria and the outer membrane in Gram-negative bacteria act as barriers blocking the entrance and access of BAC to the inner cell membrane, such cell surface modifications are expected to be beneficial to BAC resistance.

Although the above mentioned studies attempted to link specific phenotypic properties such as the active removal of BAC by efflux pumps, and reduction of the uptake, adsorption, and permeability of BAC via several types of cell surface modifications to BAC resistance, there is a lack of a comprehensive understanding of BAC-resistance mechanisms at the whole-cell level, e.g., whether or not a cell employs all these mechanisms and their relative importance in coping with high BAC concentrations. Further, it remains unclear to what specific mutations are selected by

BAC exposure. Such mutations could also serve as biomarkers for detecting BAC-resistant microorganisms as well.

To provide new insights into these questions, we exposed two strains (individual colonies) of *P. aeruginosa*, strains DP and DPB, to increasing concentrations of BAC during an adaptive evolution experiment (Fig. 1). DP and DPB strains originated from the same ancestor from a river sediment inoculum, but DP was subsequently propagated for three years in a fed-batch aerobic bioreactor with a dextrin peptone mix (DP) as a sole carbon source and never saw BAC during this incubation period, while DPB originated from a bioreactor fed with dextrin peptone plus 50mg/L BAC. Neither of the strains was able to degrade BAC in pure culture based on HPLC results; hence, biodegradation was excluded as a potential mechanism to cope with BAC exposure in the present study. We have previously sequenced the genomes of the two strains in order to evaluate whether or not the genetic adaptations, which have occurred under BAC-supplemented vs. BAC-free bioreactor conditions, have conferred increased antibiotic resistance. Briefly, we found that the two genomes differed by five point mutations, one small (9-bp) insertion, and two genomic islands, one of which appeared to be an integrative and conjugative element (ICE) likely acquired during growth in the BAC-supplemented bioreactor. At least two these genomic changes were selected by BAC exposure and conferred increased resistance to a couple antibiotics, which we will report in detail elsewhere (M. Kim, M. R. Weigand, S. Oh, J. K. Hatt, R. Krishnan, U. Tezel, S. G. Pavlostathis, and K. T. Konstantinidis, unpublished data). Here we focus on cellular adaptation to BAC-exposure by investigating mutations occurring in the genome of the two strains with different histories of BAC exposure within short periods of time (e.g., 100-200 generations) and

under increasing BAC concentrations as well as gene expression differences between BAC vs. non-BAC growth conditions.

3.3 Material and Methods

3.3.1 Adaptive evolution experiments

The BAC used in the experiments described here consisted of a 60:40 mixture of benzyldimethyldodecylammonium chloride and benzyldimethyltetradecylammonium chloride (C_{12}BDMA-Cl and C_{14}BDMA-Cl, respectively; Sigma-Aldrich). Six test tubes containing a range of BAC concentrations (e.g., a range from 200 mg/L BAC to 400 mg/L BAC) in 10 ml Luria (LB) broth medium were inoculated with a 1% aliquot of *P. aeruginosa* DP or DPB and allowed to grow for 24 hours in triplicate at 35°C, using an orbital shaker at 225rpm. The 1% aliquot originated from an overnight culture grown from a single colony of each isolate. Subsequently, 1% aliquot of the adapted population from the tube among the six tubes that showed growth at the highest minimum inhibitory concentration (MIC) was transferred to new six tubes with increasing BAC concentrations daily. The process was continued for multiple rounds until no growth was observed in any of the six tubes [BAC(+) populations]. Growth was defined as an optical density (OD) ratio >0.2 relative to the control culture grown in LB medium (no BAC). Inoculum from three BAC(+) populations (replicates) was then transferred to fresh LB medium with no BAC added [BAC(-) populations] daily, for about one hundred generations (Fig. 1A and 1B). Control experiments were conducted in parallel with the same original inoculum (i.e., ancestor strains) but no BAC added in the medium. *P. aeruginosa* DP and DPB control populations (i.e., Control_1 and Control_2) were

sampled when BAC(+) and BAC(-) populations were sampled, respectively. The number of generations of the cultures was measured every 24 hours by counting colony-forming units (CFUs). MIC was determined as the concentration at which ≥ 80% inhibition in cell growth was observed by OD 600 nm values compared to control (no BAC in the medium). The BAC antimicrobial susceptibility test was performed with the microdilution procedure (104) in LB medium. All populations were tested for their ability to biodegrade BAC using High Performance Liquid Chromatography (HPLC) as described previously (98).

3.3.2 DNA extraction, RNA extraction, and sequencing

DNA of all *P. aeruginosa* populations from the adaptive evolution experiment was extracted using the QIAamp DNA Blood Mini kit (Qiagen, Germany). DNA sequencing libraries were prepared using the Nextera XT DNA library prep kit (Illumina, San Diego, CA) and were sequenced using an Illumina MiSeq instrument (Georgia Institute of Technology). RNA of *P. aeruginosa* DPB populations from the adaptive evolution experiment was extracted as previously described (107). Briefly, cells form each *P. aeruginosa* population grown in LB or LB+BAC medium were harvested during mid-log phase, based on OD 600nm measurements. Ribosomal RNA (rRNA) was depleted from total RNA using the Ribo-Zero rRNA Removal Kit (Illumina, San Diego, CA) and cDNA libraries were constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina, San Diego, CA) according to manufacturer's instructions. All constructed cDNA libraries were sequenced using either an Illumina MiSeq instrument or an Illumina HiSEQ 2500 instrument at the Georgia Institute of Technology.

3.3.3 Bioinformatics sequence analysis

DNA of all *P. aeruginosa* populations from the adaptive evolution experiment was extracted using the QIAamp DNA Blood Mini kit (Qiagen, Germany). DNA sequencing libraries were prepared using the Nextera XT DNA library prep kit (Illumina, San Diego, CA) and were sequenced using an Illumina MiSeq instrument (Georgia Institute of Technology). RNA of *P. aeruginosa* DPB populations from the adaptive evolution experiment was extracted as previously described (107). Briefly, cells form each *P. aeruginosa* population grown in LB or LB+BAC medium were harvested during mid-log phase, based on OD 600nm measurements. Ribosomal RNA (rRNA) was depleted from total RNA using the Ribo-Zero rRNA Removal Kit (Illumina, San Diego, CA) and cDNA libraries were constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina, San Diego, CA) according to manufacturer's instructions. All constructed cDNA libraries were sequenced using either an Illumina MiSeq instrument or an Illumina HiSEQ 2500 instrument at the Georgia Institute of Technology.

3.3.4 Growth of P. aeruginosa DPB ancestor and BAC(+) populations at sub-inhibitory concentrations of BAC

To assess growth characteristics of populations, i.e., λ, lag time; and μ, maximal growth rate, the following experiment was performed. The growth medium contained sub-inhibitory concentrations of BAC, which were 100 mg/L BAC and 1000 mg/L BAC for the ancestor and BAC(+) populations, respectively. All tested inocula were first grown in LB growth media and sampled at the end of the exponential phase in order to be subsequently inoculated (diluted) in the abovementioned BAC-supplemented media to a

final concentration of 5×10^5 CFU/mL. Cell growth was performed in four replicates on 24-well plates and was measured at 600 nm using a spectrophotometer. Lag time (λ) and maximal growth rate (μ) were estimated by fitting the data with the spline fit in the grofit R package (165).

3.3.5 Data release

The isolate genome sequences were deposited at DDBJ/ENA/GenBank under the accession numbers: MTLL00000000 (*P. aeruginosa* DP) and MTLM00000000 (*P. aeruginosa* DPB), and all genomic and transcriptomic sequences of *P. aeurginosa* DPB BAC(+) and BAC(-) populations can be found under the BioProject PRJNA184698.

3.4 Results and Discussion

3.4.1 Resistance levels of P. aeruginosa exposed to increasing BAC concentrations

The initial BAC minimum inhibitory concentration (MIC) for *P. aeruginosa* DP and *P. aeruginosa* DPB was 50 mg/L BAC and 240 mg/L BAC, respectively, (Figure 3-1A) consistent with their history of BAC exposure and accumulated genomic modifications, i.e., presence of an ICE encoding a BAC efflux pump (i.e., *sugE*-A) in *P. aeruginosa* DPB and lack of the ICE in *P. aeruginosa* DP (see also gene expression results below). For the adaptive evolution experiment (Figure 3-1), six test tubes containing a range of BAC concentrations (e.g., a range from 200 mg/L BAC to 400 mg/L BAC) in Luria broth (LB) medium were inoculated with a 1% aliquot of a *P. aeruginosa* DP or DPB overnight culture and allowed to grow for 24 hours in triplicate. Subsequently, 1% aliquot of the adapted population from the tube among the six tubes

that showed growth at the highest minimum inhibitory concentration (MIC) was transferred to new six tubes with increasing BAC concentrations daily. The process was continued for multiple rounds until no growth was observed in any of the six tubes [BAC(+) populations]. All *P. aeruginosa* DP BAC(+) populations showed increased MIC for BAC after growth under increasing BAC concentrations for approximately 240 generations (DP BAC(+)_1: 920 mg/L BAC, DP BAC(+)_2: 1240 mg/L BAC, and DP BAC(+)_3: 520 mg/L BAC). *P. aeruginosa* DPB BAC(+) populations showed even higher MIC for BAC after growth with BAC for about 180 generations (DPB BAC(+)_1: 1640 mg/L BAC, DP BAC(+)_2: 1560 mg/L BAC, and DP BAC(+)_3: 1640 mg/L BAC). No biodegradation of BAC was observed by any of the populations, indicating that the BAC resistance mechanisms were physiological (e.g. differential gene expression) and/or genetic (DNA mutations).

P. aeruginosa DP and DPB BAC(+) populations were let to evolve in LB only (no BAC added) after they reached the maximum BAC concentration that they were able to withstand [thereafter referred to as BAC(-) populations; Figure 3-1] in order to test how quickly the BAC-resistance phenotype could be lost (or not). Interestingly, while all three *P. aeruginosa* DP BAC(-) populations reverted back to the similar level of the initial MIC (50 mg/L BAC) within 150 generations after transfer to BAC-free medium, two of the *P. aeruginosa* DPB BAC(-) populations maintained high BAC resistance with MIC for BAC similar to those of the DPB BAC(+) populations. The third *P. aeruginosa* DPB BAC(-) lost BAC tolerance with MIC for BAC reverting to the initial MIC (240 mg/L BAC) within about 140 generations of transferring to BAC-free medium, similar to the results observed for DP BAC(-) populations. These observed phenotypic differences

called for further genomic and transcriptomic investigations to elucidate the molecular mechanisms for BAC resistance employed by the adapted populations.

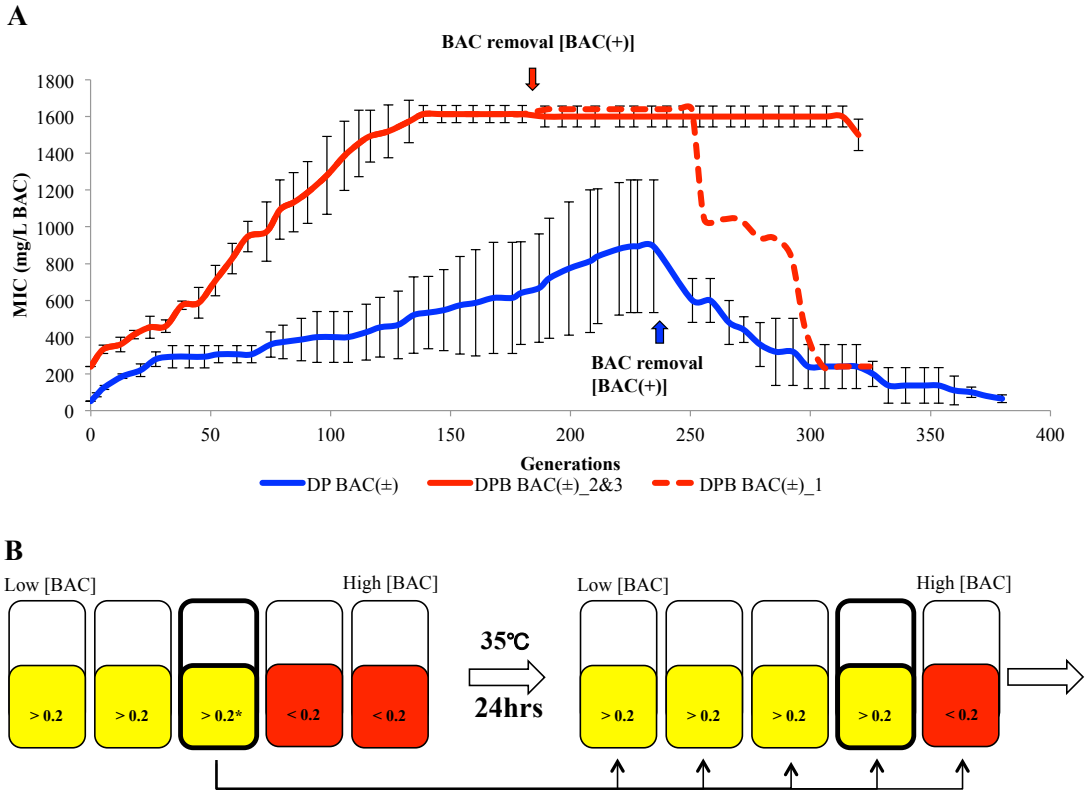


Figure 3-1. Schematic graphs of the adaptive evolution experiment. (A) Changes in minimum inhibitory concentration (MIC) of BAC. Each adaptive evolution experiment is divided into two phases; BAC-exposure, where BAC concentration was increased after each round until reaching the maximum BAC concentration in which growth was observed. And, a BAC-free phase, where BAC was removed from the growth media. Red and blue lines represent *P. aeruginosa* DPB and *P. aeruginosa* DP populations, respectively. (B) A schematic representation of the experimental design. 1% of aliquot from the culture that showed higher than 0.2 relative growth ratio compared to the control (measured by optical density; yellow test tubes) for the highest concentration of BAC tested was used as inoculum (bold black mark the inoculum tube for the example shown) for the next round with higher BAC concentrations. Red tubes represent no growth.

3.4.2 Fixed mutations in only a single gene pmrB in all BAC-adapted populations

Whole genome sequencing identified 29 SNPs and 17 DIPs (deletions, insertions and other polymorphisms) in *P. aeruginosa* DP BAC(+) and BAC(-) populations, and 22 SNPs and 16 DIPs in *P. aeruginosa* DPB BAC(+) and BAC(-) populations relative to the control population and the ancestor, respectively (**Table B 1**A, **Table B 1**B, **Table B 1**C, and **Table B 1**D). Several additional mutations identified in all *P. aeruginosa* DP and DPB populations showed lower frequencies (range: 50-80%) and were excluded from further analysis in order to focus on fixed mutations, which are more likely to underlie adaption. Based on the number of fixed mutations, the estimated spontaneous mutation rate was 6.49×10^{-9} mutations per base per generation for *P. aeruginosa* DP BAC(+) populations, which was higher (3.85-fold, $P = 0.016$, Student's t-test) than that of the *P. aeruginosa* DP BAC(-) populations (1.68×10^{-9} mutations per base per generation). On the other hand, the estimated spontaneous mutation rate for *P. aeruginosa* DPB BAC(+) populations was 7.34-fold higher than that of the *P. aeruginosa* DPB BAC(-) populations (7.72×10^{-9} vs. 1.05×10^{-9}), which was significant ($P < 0.01$, Student's t-test). Therefore, higher mutation rates were observed during BAC-adaptation compared to growth on BAC-free medium for both *P. aeruginosa* isolates, and the change in mutation rate was higher for DPB populations, which withstood higher BAC concentrations.

Notably, all DP BAC(+)/BAC(-) and DPB BAC(+)/BAC(-) had fixed mutations only in a single gene in common, *pmrB*, albeit at different locations within the gene. Mutations in *pmrB* are known to confer polymyxin antibiotic resistance and the level of

polymyxin antibiotic resistance depends on the type of mutation in *pmrB* (e.g., location and combinations of mutation) (137). Apparently, these mutations were selected by BAC exposure and represent a common mechanism to cope with BAC toxicity, at least for *P. aeruginosa* (Note that polymixin is also a membrane-disrupting agent, similar to BAC) (139, 140). In particular, mutations in *pmrB* are known to lead to constitutive activation of the *pmrA* regulon, and expression of *pmrA* leads to the expression of *arnBCADTEF*, which is responsible for the addition of 4-amino-L-arabinose (L-Ara4N) to the phosphate groups of lipid A, resulting in reduction of the net negative charge on the outer membrane (136-138). The reduction in net negative charge could increase bacterial BAC tolerance by reducing adsorption of the positively-charged BAC molecules to the cell wall.

The loss of BAC resistance in DP BAC(-) populations was attributable to the acquisition of different mutations in other parts of *pmrAB* or other regulatory genes selected by BAC(-) growth conditions (**Table B 1**C and **Table B 1**E). In particular, *P. aeruginosa* DP BAC(-)_2 and DP BAC(-)_3 subpopulations had additional mutations at different locations in *pmrB* and *pmrA* compared to those in *P. aeruginosa* DP BAC(+) populations (e.g., SNPs in *pmrB* in >50% of the population), and *P. aeruginosa* DP BAC(-)_1 had a 15bp insertion mutation in the quorum sensing regulator gene *lasR*, which controls more than 300 genes including various efflux pump genes (e.g., MFS, and RND family) (166, 167) (**Table B 1**C). These results may explain the loss of BAC resistance in all DP BAC(-) populations, although direct experimental verification of these results will be subject of future research.

Other than *pmrB*, no consistent mutations were identified in a functional gene that was shared by all of the triplicate populations of either DP or DPB strains. The latter

findings indicated that additional adaptations to BAC exposure might have been at the level of cell physiology and/or gene regulation. Therefore, we conducted transcriptomic analysis for *P. aeruginosa* DPB BAC(+) and BAC(-) populations both under BAC-free and sub-inhibitory BAC concentrations in order to identify molecular mechanisms for BAC adaptation.

3.4.3 Gene expression under BAC-exposed conditions

To further investigate the molecular mechanisms for BAC resistance in *P. aeruginosa* DPB, we first conducted pairwise comparisons of gene expression profiles determined by RNA-seq of the ancestral population and its evolved control population (under BAC-free medium, LB) of *P. aeruginosa* DPB when grown with BAC-free versus BAC-exposed media (i.e., sub-inhibitory concentration of BAC were used). These pairwise comparisons identified 1071 differentially expressed genes in the ancestor and 1474 differentially expressed genes in the control out of the total 6038 predicted genes present in the genome (Table B 2). Notably, expression of efflux pump genes (i.e., *sugE*-A, *sugE*-B, ATP-binding cassette (ABC) family, and a resistance nodulation division (RND) family member) in the ICE increased in both the ancestral and control populations under the BAC-exposed vs. BAC-free conditions (Figure 3-2). The expression level of *sugE*-A, in particular, was slightly increased in the ancestor (1.4 fold higher), but not statistically significantly (padj $>$ 0.01), and increased significantly in the control population (4 fold higher with padj $<$ 0.01) under BAC-exposed conditions. *sugE*-A has also been shown previously to confer increased BAC resistance in *Enterobacter cloacae*

(132). Collectively, these results suggested that the ICE conferred BAC-resistance based on the presence of *sugE*-A and other transport proteins.

Furthermore, three protein sequences of the ICE encode for a RND transporter that showed > 95% amino acid sequence identity with the AcrAB-TolC efflux pump in *Salmonella enterica* subsp. enterica serovar Typhimurium, except for AcrA, which showed 25% amino acid sequence identity. BAC resistance mutants of *S.* Typhimurium showed increased expression of *arcB* and deletion mutants for *acrAB* and *tolC* showed increased susceptibility to BAC (168). Similar to the *sugE*-A genes mentioned above, the ancestor showed slightly increased, but not significant, expression of *acrB* and *tolC* (1.7 and 2.3 fold higher, with padj > 0.01, respectively), and the control population showed significantly increased expression of all these genes (Figure 3-2). These results suggested that the efflux pump genes might also be responsible, at least in part, for BAC resistance.

Further, the different degrees of overexpression of these genes under BAC-exposed conditions in the control and the ancestor population compared to the BAC-free conditions maybe be attributable to the exposure history of the populations. The control grew for ~330 generations in LB medium before the MIC for BAC test was performed, while the ancestor experienced growth for ~3 years in a BAC-amended bioreactor before the test. Thus, the ancestor was perhaps (more) pre-adapted to BAC presence, which likely accounted, at least in part, for lower induction of the efflux pump genes in the ICE relative to BAC-free conditions. Consistent with this explanation, background ICE gene expression in the ancestor was higher, by 2.8 fold, on average than that in the control under BAC-free conditions ($P < 0.01$, Student's t-test), but no such significant difference was observed under BAC-exposed conditions. It is also important to note that the ICE

was likely present in the highly complex microbial community of the starting river sediment inoculum given that *P. aeruginosa* DPB, who encoded the ICE, and DP isolates represented descendants of the same ancestor in the original inoculum based on sequence analysis (e.g., it is less likely that the ICE element was brought into the bioreactors through aerial transport in the meanwhile). However, the exact genomic context that the ICE resided in the original inoculum remains unknown.

In order to further corroborate these findings, and provide insights into the specific adaptations that enable evolved *P. aeruginosa* DPB strains to withstand high concentrations of BAC, we conducted pairwise comparisons of gene expression between the ancestral population and the *P. aeruginosa* DPB BAC(+) populations, under both BAC-free (Table B 3 and Table B 4) and BAC-exposed conditions (Table B 5 and Table B 6). Differentially expressed genes in the evolved control population versus the ancestor were considered to be the result of stochastic processes or selection by the growth conditions (e.g., "bottle effect"), and these genes were removed from the comparison of the ancestor versus the *P. aeruginosa* DPB BAC(+) as also performed in several previous studies, e.g., (169). In the pairwise comparison between the ancestor and *P. aeruginosa* DPB BAC(+) populations under BAC-exposed conditions, we were able to identify 303 differentially expressed genes that were shared among the three *P. aeruginosa* DPB BAC(+) populations (Table B 5), while 381, 272 and 710 differentially expressed genes were specific to *P. aeruginosa* DPB BAC(+)_1, *P. aeruginosa* DPB BAC(+)_2, and *P. aeruginosa* DPB BAC(+)_3, respectively (Table B 6). This number of genes was significantly larger the one form a similar comparison under BAC-free condition (identified 61 differentially expressed genes in common), indicating that important

physiological adaptations to BAC exposure have taken place in the replicate *P. aeruginosa* DPB BAC(+) populations and underlay their (shared) high MIC levels for BAC. Consistent with the latter interpretations, the total number of differentially expressed genes (n=373) found in the comparison between *P. aeurginosa* DPB BAC(+)_1 and *P. aeurginosa* DPB BAC(-)_1, which lost BAC resistance, under BAC-exposed conditions was higher than the number of such genes (n=16) found in the comparison between *P. aeruginosa* DPB BAC(+)_2 and *P. aeruginosa* DPB BAC(-)_2, which maintained MIC for BAC, under BAC-exposed conditions (Table B 7 and Table B 8). Further, no differentially expressed genes were found between the replicate populations, e.g., three replicates of BAC(+)_1 population, under the same conditions (i.e., BAC-free or BAC-amended), indicating that the differences identified by our RNA-seq analysis above reflected largely biologically relevant adaptations and not merely higher noise of RNA versus DNA datasets and protocols. Therefore, our subsequent analysis was focused on these 303 genes differentially expressed in *P. aeruginosa* DPB BAC(+) populations compared to the ancestor under BAC-exposed conditions.

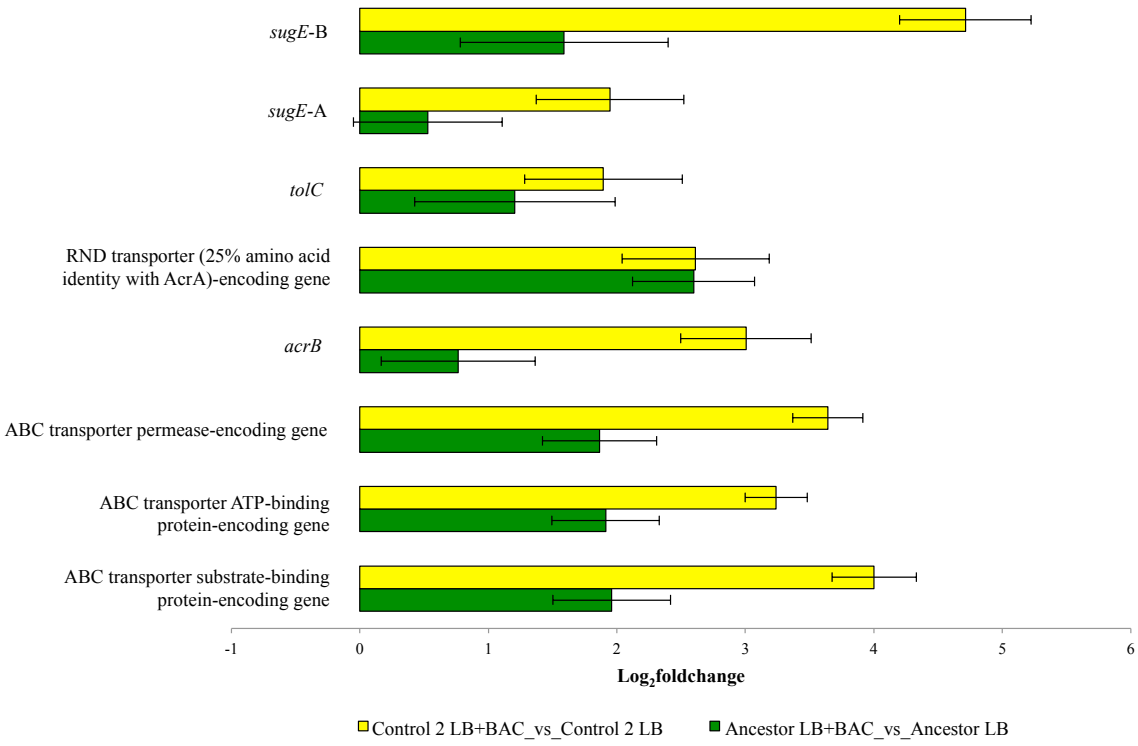


Figure 3-2. Overexpression of efflux pump genes present in the ICE when BAC is supplemented in the growth media. Green bars represent comparisons between *P.aeruginosa* DPB ancestor in LB growth media versus LB+BAC media (AL_vs_AB); yellow show comparisons between *P. aeruginosa* DPB_Control_2 in LB vs. LB+BAC (C2L_vs_C2B). AL: *P.aeruginosa* DPB ancestor in LB growth media; AB: *P.aeruginosa* DPB ancestor in LB+BAC media; C2L: *P. aeruginosa* DPB_Control_2 in LB media; C2B: *P. aeruginosa* DPB_Control_2 in LB+BAC media.

3.4.4 Restricted growth as an adaptation to high BAC concentrations

With a sub-inhibitory concentration of BAC in the growth medium, expression of phosphorus utilization related genes significantly decreased in the evolved DPB BAC(+) populations compared to the ancestor (Figure 3-3A and Table B 5). It is unlikely that these expression patterns were attributable to the lower phosphorus concentration or different phosphorus species in the LB medium because the control population, which

was propagated on LB, did not show similar gene expression patterns (note also that expression differences in control vs. ancestor were removed from further analysis in order to avoid such effects of LB growth medium, as noted above). Significantly decreased expression (-2.24 $\log_2$ fold change, on average, padj < 0.01) of the phosphate-specific transport (*pst*) operon (*pstSCAB* and *phoU*) and the *phoB* gene, encoding a transcription activator of the two component regulatory system *phoBR,* was observed in all *P. aeruginosa* DPB BAC(+) populations. *P. aeruginosa* possesses a *pst* system (high affinity phosphate transport) which is activated by phosphate limitation (170) leading to phosphate taxis, and deletion of the *pstCAB* and *phoU* genes results in restrain of phosphate taxis even in conditions of phosphate excess (171). Therefore, decreased expression of *pst* operon suggested lower uptake of phosphate. Furthermore, significantly decreased expression of phosphonate utilization pathway (-2.69 $\log_2$ fold change, on average, padj < 0.01) was observed in all *P. aeruginosa* DPB BAC(+) populations, especially the ATP-binding protein of the phosphonate ABC transport system *phnC*, ABC transporter substrate-binding protein *phnD*, phosphonates transporter *phnE*, and carbon-phosphorus lyase complex subunit *phnJ*. Phosphonates are P compounds (valence +3) with a C-P bond instead of the more common C-O-P bond (valence +5 of P) found in phosphate esters, so the strength of C-P bond makes phosphonates resistant to chemical and enzymatic hydrolysis (172). Some microorganisms are known to be able to utilize phosphonates as a P source and phosphonates utilization pathways have been reported with detailed genetic and biochemical evidence (173, 174). Therefore, the lower expression of these genes was also consistent with decreased phosphorus utilization as an effect of exposure to high BAC concentrations.

Adaptation to high concentrations of BAC resulted in additional significant changes, mostly a decrease in expression of genes related to various metabolic pathways such as fatty acid biosynthesis, the nitrogen cycle and energy metabolism in all *P. aeurginosa* DPB BAC(+) populations compared to the ancestor (Figure 3-3A). Firstly, decreased expression of acetyl-CoA carboxylase biotin carboxyl carrier protein subunit, acetyl-CoA carboxylase biotin carboxylase subunit, and 3-oxoacyl-ACP synthase III, which are involved in the fatty acid biosynthetic process (-2.17 $\log_2$ fold change, on average, padj < 0.01), and decreased expression of acyl-CoA dehydrogenase (-1.33 $\log_2$ fold change, on average, padj < 0.01), which is the initial step of fatty acid beta-oxidation, were observed. Secondly, the expressions of almost all genes related to energy metabolism (i.e., the electron transport system) were decreased. For instance, the expression of genes involved in an electron transfer chain such as those encoding the cbb3-type cytochrome c oxidase subunits (i.e., *ccoN*, *ccoO*, *ccoQ*, and *ccoP*), the cytochrome c biogenesis gene *ccsA*, and the cytochrome c4, 2-nonaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase which is part of ubiquinone (cofactor) biosynthetic pathway, and ferredoxin decreased by 2.43 $\log_2$ fold change, on average (padj < 0.01), relative to the ancestor. Furthermore, significantly decreased expression of *roxR* (encoding a response regulator of RoxSR two-component regulatory system) and cyanide insensitive terminal oxidases (i.e., *cioA* and *cioB*) were observed in all *P. aeruginosa* DPB BAC(+) populations (-1.71 $\log_2$ fold change, on average, padj < 0.01). This result was consistent with previous findings that RoxR is a direct positive transcriptional regulator of a cyanide-insensitive oxidase that is part of the electron transport system in *P. aeuriginosa* (175). Finally, we observed significantly reduced

expression of ribonucleoside-diphosphate reductase (-1.78 $\log_2$ fold change, on average, padj < 0.01), which is responsible for the production of the precursors necessary for DNA synthesis. Altogether, reduced expression of genes related to cell energy metabolism, various metabolic pathways, and decreased uptake and utilization of phosphorus suggested that evolved *P. aeruginosa* DPB BAC(+) populations adapted to high concentrations of BAC might alter their growth characteristics (e.g., lag time or growth rate).

Reduction of the growth rate is known to be beneficial for bacterial populations exposed to antibiotics such as beta-lactams, which are cell-envelope disrupting agents like BACs, and other stress exposed conditions (176, 177). Therefore, slow growth rate might be beneficial for the high concentration of BAC adaptation. Consistent with these interpretations, we noted a slower growth rate in all DPB BAC(+) populations growing in 1000mg/L BAC relative to that of the ancestor grown in the presence of 100mg/L BAC ($P < 0.01$, Student's t-test) (Table 3-1). Furthermore, the average total cell count of the ancestor during the stationary phase under BAC-exposed conditions (100mg/L BAC) was 3.87*10^7 CFU/mL, which was 2.5 fold higher than that (1.52*10^7 CFU/mL) of *P. aeruginosa* DPB BAC(+) populations under BAC-exposed conditions (1000mg/L BAC) ($P < 0.05$, Student's t-test).

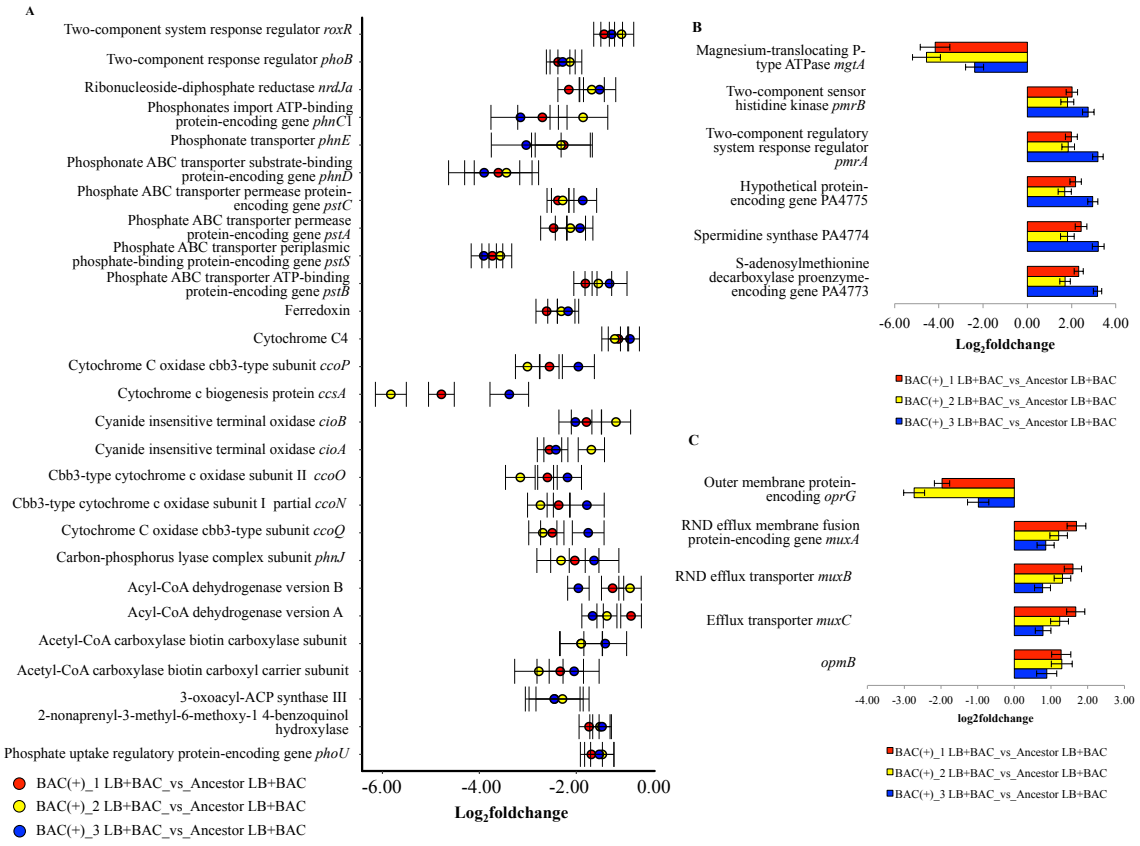


Figure 3-3. Gene expression changes in BAC-adapted *P. aeruginosa* DPB populations. Average $\log_2$ fold change values of three replicate for each population and ancestor were used. (A) Decreased expression of genes related to growth in *P. aeruginosa* DPB BAC(+) populations versus their ancestor in LB+BAC media (AB_vs_BAC(+)). (B) Overexpression of spermidine synthesis genes and *pmrAB*, and decreased expression of Mg^{2+} transport ATPase. (C) Overexpression of *muxABC-opmB* RND efflux pump and decreased expression of *oprG* porin, for the same comparison. AB: *P.aeruginosa* DPB ancestor in LB+BAC media; BAC(+)_1B: *P. aeruginosa* DPB BAC(+)_1 population in LB+BAC media; BAC(+)_2B: *P. aeruginosa* DPB BAC(+)_1 population in LB+BAC media; BAC(+)_3B: *P. aeruginosa* DPB BAC(+)_3 population in LB+BAC media.

Table 3-1. Growth characteristic parameters in *P. aeruginosa* DPB ancestor and BAC(+) populations at sub-inhibitory concentrations of

BAC. See Experimental procedures for details on how cells were grown and growth rate was measured.

	Ancestor with BAC	**BAC(+)_1 with BAC**	**BAC(+)_2 with BAC**	**BAC(+)_3 with BAC**
Maximal growth rate (OD 600nm/hour)	0.15 ± 0.01	0.11 ± 0.004	0.06 ± 0.01	0.13 ± 0.04
Lag time (hours)	4.36 ± 0.44	6.69 ± 0.24	2.67 ± 0.37	6.4 ± 0.33

3.4.5 Stress-related genes

We also examined additional genes that were suspected to be involved in BAC-resistance based on the (known) mode of action of BAC. Exposure to sub-lethal concentrations of cetyltrimethylammonium bromide, another member of QAC family, induces oxidative stress by generating reactive oxygen species (ROS, e.g., superoxide and hydrogen peroxide) in *E. coli* cells (178). Accordingly, bacteria subjected to BAC are known to induce genes for ROS-scavenging antioxidants such as catalase and alkyl hydroperoxide reductase (179, 180). The expression of the gene encoding universal stress protein A (UspA) is also elevated in response to various kinds of stress conditions including oxidant exposure and in the presence of polymyxin or other antibiotics (181). However, when *P. aeruginosa* DPB BAC(+) populations were exposed to sub-inhibitory concentrations of BAC, the expression of alkyl hydroperoxide reductase, catalase, and *uspA* decreased by 1.49 of average $\log_2$ fold (padj < 0.01) relative to the ancestor (Table B 5), indicating that expression of these stress-related genes does not represent a major physiological adaptation to BAC exposure, at least in the *P. aeruginosa* strains. On the other hand, we observed elevated expression of stress-induced gene *ygiW* (+2.64 $\log_2$ fold

change, on average, padj < 0.01) in all *P. aeruginosa* DPB BAC(+) populations (Table B 5). The *E. coli ygiW* mutant is sensitive to hydrogen peroxide and produces elevated levels of biofilm with the expression of flagellar biosynthesis genes (182); *ygiW* is also known to be induced by hydrogen peroxide (183). Although the mechanisms by which *ygiW* protects cells from hydrogen peroxide remain unknown, the increased expression of *ygiW* in all *P. aeruginosa* DPB BAC(+) populations is consistent with the previous studies that QAC induce oxidative stress by generating ROS (178). Overall, these results indicate that protection from the oxidative stress induced by exposure to a sub-inhibitory concentration of BAC might not be the primary BAC resistance mechanism that allows *P. aeruginosa* DPB BAC(+) populations to withstand high concentrations of BAC compared to their non-adapted ancestor.

3.4.6 Spermidine

Extracellular DNA can function as a nutrient source, a biofilm component and as a cation chelator, which can result in growth inhibition and thus, has implications for antibiotic resistance (184). Because DNA can sequester divalent metal cations, exposure to high levels of extracellular DNA induces the Mg^{2+}-responsive PhoPQ and PmrAB two-component systems, which control many genes required for virulence and antimicrobial peptide resistance (184). Exposure to sub-lethal concentrations of extracellular DNA and membrane-targeting antibiotics induces expression of genes *PA4773* to *4775* of the *P. aeruginosa* type strain, which are putatively annotated as spermidine and other polyamine synthesis genes. Polyamines are produced under Mg^{2+}-limiting conditions and can act as an organic polycation to bind lipopolysaccharide (LPS). Thus, these polyamines can stabilize and protect the outer membrane against

antibiotic (i.e., polymyxin B) and oxidative damage (i.e., hydrogen peroxide) (185). Interestingly, both expression of spermidine synthase gene operon, which is almost identical to genes *PA4773* to *4775*, and expression of *pmrAB* operon genes, which are located downstream of the spermidine synthase operon (i.e., *PA4776* and *PA4777*), were significantly increased by a 2.39 and a 2.27 $\log_2$ fold change, on average (padj < 0.01), respectively, in all *P. aeruginosa* DPB BAC(+) compared to the ancestor (Figure 3-3B and Table B 5). Furthermore, significantly decreased expression of Mg^{2+} transport ATPase (*mgtA*) was also observed (-3.71 $\log_2$ fold change, on average, padj < 0.01) (Figure 3-3B and Table B 5). Therefore, these results indicate that a chloride ion in BAC might act as a cation (e.g., Mg^{2+}) chelator, destabilizing the cellular membrane, and *P. aeruginosa* DPB BAC(+) populations evolved to express more spermidine synthase genes when exposed to BAC conditions in order to protect cells from either membrane damaging BAC and/or oxidative stress possibly induced by BAC exposure.

3.4.7 Efflux pumps and porins

In addition to the ICE-encoded efflux pumps genes mentioned above, we also examined the chromosomal encoded ones. MuxABC-OpmB is an RND-type multidrug efflux pump that has been reported to confer resistance to novobiocin, aztreonam, macrolides, and tetracycline (186), and inactivation of the MuxABC-OpmB transporter leads to ampicillin and carbenicillin susceptibility (187). Moreover, the *muxABC-opmB* genes expressed by metabolically active subpopulation in *P. aeruginosa* biofilm when exposed to colistin (polymyxin E) was necessary for colistin resistance but did not affect resistance to the chelator EDTA, the detergent SDS, or chlorhexidine (188). A modest increased expression of the *muxABC-opmB* genes (+1.21 $\log_2$ fold change, on average,

padj < 0.01) (Figure 3-3C and Table B 5) indicated that BAC may also be a substrate for this efflux pump. However, the latter finding awaits experimental verification because none of the previous studies of the *muxABC-opmB* genes was focused on BAC. OprG is one of the major outer membrane proteins (e.g., OprD, OprE, OprF, and OprH) in *P. aeruginosa* (189). The function of OprG has not been clearly elucidated and several studies reported that decreased expression of *oprG* was related to increased resistance to antibiotics such as tetracycline and kanamycin (190). However, other studies reported that expression of *oprG* under iron-rich anaerobic conditions was increased but OprG was not involved in iron or antibiotic uptake (191) and interestingly OprG might mediate the diffusion of small hydrophobic molecules across the outer membrane based on structural analysis (192). Because the first step of mode of action for BAC is adsorption and penetration of the BAC into the cell wall and BAC has a long hydrophobic alkyl chain, which can be a substrate for OprG, decreased expression of the *oprG* gene by a 1.89 $\log_2$ fold change, on average (padj < 0.01) in all *P. aeruginosa* DPB BAC(+) populations compared to the ancestor at sub-inhibitory concentration of BAC might be beneficial to prevent BAC from entering into the cells (Figure 3-3C and Table B 5). However, further experimental tests are necessary to test these hypotheses in order to clarify the mechanisms of BAC resistance.

3.5 Conclusions and outlook

It should be noted that BAC are often combined with other chemicals (e.g., excipients) in typical disinfectant formulations, which could increase BAC potency (193); no excipients were used in our study. Nonetheless, experimental adaptation to increasing concentrations of BAC rendered *P. aeruginosa* resistant to high BAC

concentrations, ranging from 480 mg/L BAC to 1600 mg/L BAC, which are comparable to, or even higher, than those used in practice for disinfection, e.g., typically between 400-500 mg/L (146). Given the high stability of BAC in the environment and several outbreaks linked to contaminated QAC-based antiseptic solutions, these results underscored the need for further investigation of the mechanisms of BAC resistance.

Our genomic and transcriptomic analysis on BAC-evolved *P. aeruginosa* suggested that BAC resistance was both at the level of cell physiology (e.g., gene regulation) and genomic adaptations (e.g., fixed mutations). After 3 years of exposure to a sub-inhibitory concentration of BAC within the DPB bioreactor, *P. aeruginosa* acquired an ICE encoding several efflux pump proteins that were likely responsible for BAC resistance based on bioinformatics gene function annotation. In the present study, we showed that these genes were also overexpressed under BAC-exposed conditions compared to BAC-free conditions (e.g., Figure 3-2), further corroborating the bioinformatics predictions that the genes are related to BAC resistance. Moreover, the physiological adaptations to high BAC concentrations appeared to include a combination of reduction of phosphorus utilization/uptake and energy production (e.g., electron transfer chains), slower growth in general, and altered expression of efflux pump genes and porins. These extensive and multi-facet differences at the transcriptome level contrasted with those at the DNA level, which revealed only a single gene in common that had undergone mutations during BAC exposure (*pmrB*). Specifically, after experimental adaptation to high concentrations of BAC, fixed mutations in *pmrB* allowed *P. aeruginosa* to be resistant to BAC via expression of genes related to stabilization of cell surface charge (e.g., 4-amino-L-arabinose and spermidine synthesis). Therefore,

several independent mechanisms were employed by evolved *P. aeruginosa* populations to tolerate ~7 fold higher BAC concentration relative to its ancestral population.

Our results revealed that reduction of growth and stabilization of cell surface charge represented two of the major BAC resistance mechanisms. This finding suggested that highly-sensitive, culture-independent detection methods such as qPCR could be used to monitor the microbial contamination of disinfectant manufacturing facilities or hospital disinfectant stocks in order to prevent outbreaks associated with contaminated stocks or storage (35, 157). In particular, mutations in *pmrB* can be used as potential biomarkers to detect BAC-resistant organisms, especially *P. aeruginosa*. Furthermore, because our results suggested that stabilizing cell surface charge by the production of cationic polyamines represents another major BAC resistance mechanism, adjusting (e.g., lowering) the pH of the water in the antiseptic processing lines could be helpful in removing any remaining BAC-resistance pathogens. However, adjusting the pH might not be always easy or even possible. Finally, BAC evolved *P. aeruginosa* showed changes in the expression of several virulence factors (e.g., Type III secretion systems and genes related to motility) under both BAC-free and BAC-exposed conditions (Table B 3, Table B 4, Table B 5, and Table B 6). These results implied that exposure (and resistance) to BAC of pathogens may also alter their virulence, e.g., decreased expression of Type III secretion systems and motility genes in BAC adapted populations, or increased expression of pyoverdine biosynthesis genes (Table B 3, Table B 4, Table B 5, and Table B 6), which warrants further investigation.

It is important to note that our experimental design (e.g., semi-continuously fed, mixed batch culture) represents a good model for engineered biological wastewater

treatment systems used in poultry meat, pharmaceutical processing facilities). For instance, BAC often accumulate in such facilities due to the continuous input and adsorption to biomass (194) and thus, our experimental design based on increasing BAC concentrations (Figure 3-1B) simulate this scenario well. Further, our results showed that the triplicate BAC adapted *P. aeruginosa* DPB populations were able to maintain their high MIC even after BAC was removed from the feed for at least one hundred generations. Hence, our findings are -in general- realistic with respect to practice.

3.6 Acknowledgements

This work was supported by US National Science Foundation awards 0967130 to SGP and KTK and 1241046 to KTK.

CHAPTER 4 METAGENOMIC INSIGHTS INTO THE EFFECTS OF IMPROVED SANITATION ON ENTERIC GUT INFECTIONS AND THE MICROBIOME OF CHILDREN IN MOZAMBIQUE

4.1 Abstract

Improper or lack of sanitary conditions can enhance the infection of enteric pathogens, resulting in illness and even death of children in low-income countries. By employing metagenomic analysis of stool samples, we examined associations between several environmental variables, including sanitation interventions, and gut microbiome composition in 122 children under 48 months old in a low-income, urban neighborhood setting in Maputo, Mozambique. While age was the most influential variable on microbial community composition differences among the children sampled, antibiotic resistance gene (ARG) and virulence factor (VF) abundances as well as other variables including sanitation intervention, diarrhea status, compound sanitary index, and compound level population density also contributed significantly to the differences observed (P value of < 0.05, PERMANOVA). Especially, *Veillonella parvula*, an opportunistic pathogen, was higher in the control group (no sanitation) and diarrhea group compared to the sanitation intervention and the non-diarrhea (no disease symptoms) groups. We were able to recover several genomes representing known as well as novel taxa that were differently abundant between the sanitation intervention group

and the control group and thus, should be subjects of future research. Collectively, these results suggested that sanitation interventions might induce small but significant differences in abundance of at least several microbial species, including *V. parvula*, and expanded the genomic diversity of species present in the children microbiome. However, pathogen load in the metagenomes analyzed remain high in children with intervention compared to their matched controls, indicating alternative routes of infection that remain uncontrollable.

4.2 Introduction

Diarrhea causes 1.7 million annual deaths, mostly to children living in low-income countries where basic sanitation conditions do not exist (43, 48, 195). Several distinct species of bacteria, viruses, and protozoa have been identified as the leading causative agents of diarrhea (85, 196). However, the impact of enteric infections is likely underestimated because of the presence of asymptomatic individuals (85, 196, 197). These symptomatic and asymptomatic infections have major effects on (lower) cognitive and physical growth in early childhood, leading often to death, and have been generally reported to be associated with poor water, sanitation and hygiene (WASH) conditions (198, 199).

Several recent studies have reported no significant associations between improved WASH conditions, especially sanitation interventions, and enteric infections or health conditions such as the incidence of diarrhea (81-84). However, the effects of WASH conditions on enteric infections can be complex and not easy to quantify due, for instance, to lack of precision in WASH outcome measures, including self-reported (as opposed to objectively measured) diarrhea and comparison of data from various settings

such as rural vs. urban areas (75, 80, 200, 201). Self-reported diarrhea does not account for asymptomatic infection either (202-204). Furthermore, the majority of studies regarding sanitation have focused on rural regions while only a few studies have attempted to understand the effect of sanitation on human health in rapidly growing urban areas (48, 75, 205). Densely populated urban areas without proper sanitary infrastructure can facilitate fecal contamination and enteric pathogen transmission within the community, which can increase the risk of exposure to pathogens and thus, render sanitation interventions ineffective (47, 48). Therefore, there is a need to investigate the association between sanitation and human health in urban areas that lack adequate sanitation facilities.

Furthermore, many studies in the WASH research field have employed a standardized microbiology protocol from the Etiology, Risk Factors, and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development Project (MAL-ED) for sample collection and analysis (85, 86). Specifically, this protocol employs culture, microscopy, immunoassay, and PCR for the detection of a specific list of pathogens. In addition to this, a number of other recent WASH studies used multiplex PCR assay to determine the presence or absence of representative pathogens (48, 87, 88). These conventional procedures for the detection and identification of various pathogens are time-consuming and labor intensive, and the PCR based approaches can even produce false-positive signal due to primer hybridization with non-target DNA/RNA (206). Further, the primers used in the PCR assays are designed based on available genomes and thus, cannot capture novel or divergent at the sequence level pathogens. Contrasting with these conventional methods, metagenomic

sequencing represents a more unbiased approach that can theoretically detect and genotype all pathogens present in a sample and can uncover unknown pathogens or commensals (207). Preliminary analysis of a small sample size has shown that the gut microbiomes of non-westernized populations, including Africans, show profound differences compared to those of westernized populations (208, 209). Thus, it is important to apply metagenomic approaches to examine the association between sanitation and gut microbiome, including but not limited to the presence or absence and relative abundance of enteric pathogens in the gut microbiomes of African children, which are currently underrepresented in the public databases.

In the context of Maputo Sanitation (MapSan) trial in urban neighborhoods of Maputo, Mozambique (64), which employed pour-flush latrines (to septic tank) sanitation intervention shared by multiple households, we investigated the effect of various variables, including sanitation conditions, on the gut microbiome of children under 48 months old in a low-income.

4.3 Materials and Methods

4.3.1 The Maputo Sanitation (MapSan) study design

The MapSan project, a controlled before-and-after study of the impact of an urban sanitation intervention on children health in Maputo, Mozambique, was conducted between February 2015 and September 2018 (64). The sanitation intervention, built by the non-profit implementation agency Water and Sanitation for the Urban Poor, involved shared private latrines used by a minimum of two households, with pour-flush toilets to septic tanks (non-sewerage). The number of latrine cabins depends on the number of

beneficiaries in the compound (grouping of houses sharing sanitation and outdoor living space). Compounds with ≤20 people were provided a single cabin "Shared Latrine," and compounds with >20 people received a "Communal Sanitation Block" (CSB) with one latrine cabin per 20 people residing in the compound. The CSB interventions also included facilities such as laundry and washing units, rainwater and municipal water storage, and handwashing stations. Control sites, which did not receive upgraded sanitation, were enrolled concurrently with intervention sites. Selection criteria for intervention and control sites have been described in detail elsewhere (1, 2). Briefly, all study sites must have had (i) shared sanitation in poor condition (e.g., pit latrines or pit latrines with slabs or non-function pour flush latrines), (ii) a piped water supply nearby, (iii) a minimum number of beneficiaries, (iv) stated demand for improved sanitation, and (v) an eligible child. When possible, control sites were selected to have a similar number of residents as concurrently enrolled intervention sites. The criteria for enrollment of children were (i) age range between 1 and 48 months at the time of enrollment, (ii) available written informed consent by child's guardian, and (iii) met site criteria described above. The study protocol was approved by the Comite´ Nacional de Bioe´tica para a Sau´de (CNBS), Ministe´rio da Sau´de (333/CNBS/14), the Ethics Committee of the London School of Tropical Medicine and Hygiene (reference # 8345), and the Institutional Review Board of the Georgia Institute of Technology (protocol # H15160). The associated MapSan trial has been registered at ClinicalTrials.gov (NCT02362932).

In total, we sequenced 122 stool samples from children in three age groups, i.e., 1-11 months, 12-23 months and 24-48 months. The samples were categorized into four different groups: baseline control, baseline intervention, midline control and midline

intervention (Table 4-1A). The first group of samples selected for metagenome sequencing in Spring 2017 were all aged 7-14 months old. To select these samples, the following procedure was used. First, we identified as many samples as possible from the midline phase that were between 8 and 12 months old, had stool available in enough volume for metagenomics, and were under sanitation intervention. These were matched to samples from children with no intervention. Matching criteria were age (+/- 1 month) and sex. In total, we ended up with 26 intervention (13 from baseline, meaning first sample taken before sanitation is applied, and 13 from midline phase, after sanitation is applied) and 16 control (8 from baseline from controls that no sanitation was applied, and 8 from midline phase) for a total of 42 samples. For the second group of samples selected for sequencing in December 2017, we chose samples from two different groups of children: (i) 40 additional samples from children 7-14 months at midline phase, equally split between intervention and control, and (ii) 40 samples from children aged 2-3 years (i.e., older children, being under intervention for 12 month) at midline phase, equally split between intervention and control. Most of sub-group (i) were also subjected previously to AMR qPCR (35/40) for comparison to the qPCR array results. The other 5 for this age sub-group were chosen randomly after imposing certain practical criteria, i.e., making sure that equal numbers of intervention and control samples were used (20 each), and that there was enough stool material available for analysis. 12 of the sub-group (ii) samples (2-3 year old; out of 40, in total) were chosen because we had included the matched baseline sample in the first group of samples described above (1st paragraph) and these represent samples from children under sanitation that can be compared to the samples from the same children before sanitation. The remaining 28 were randomly selected after

taking into consideration samples with enough stool material for the analysis and achievement of balance of intervention vs control (20 of each).

We defined diarrhea as ≥ 3 loose or liquid stools in a 24-hour period or any stools with blood (210). Metadata used in this study are presented in Table 4-1B. Breastfeeding was present in all 1-11 months old samples. Thus, breastfeeding was not independent of age, and we did not consider breastfeeding as an independent variable.

4.3.2 DNA extraction and sequencing

Approximately 0.1 g of each fecal sample was taken for total DNA extraction and processed with the standard Manual of Procedures (MoP) suggested by the Human Microbiome Project (HMP- http://hmpdacc.org/resources/tools_protocols.php). DNA quantification was achieved using Quibit dsDNA HS Assay Kit (Thermo Fisher Scientific). DNA libraries were prepared using the Nextera XT DNA library prep kit (Illumina, San Diego, CA) and were sequenced using an Illumina HiSeq 2500 instrument (Georgia Institute of Technology).

4.3.3 Bioinformatics data analysis

Metagenomic reads were quality-trimmed with SolexaQA, using phred quality cut-off of 20. Reads shorter than 50 bps after trimming were discarded (108). Trimmed reads were then filtered by BMTagger to identify human reads (211). Non-human reads were then used for the rest of the analyses to investigate the gut microbiome. The estimated abundance-weighted average coverage of each metagenomic dataset was calculated by Nonpareil version 3.20 with default parameters (212). Mash, a tool that uses the MinHash dimensionality reduction method to compare sample-to-sample kmer

composition (*i.e.*, β-diversity) (213), was used to calculate pairwise mutation distance between whole metagenomic datasets with a kmer=25 option. Microbial community taxonomic composition was assessed based on clade-specific marker genes using MetaPhlAn2 v2.0 (214). To assess the abundance of antibiotic resistance genes (ARGs), deepARG (215), which identifies ARG-encoding metagenomic reads after searching them against an ARG database was used. Blastn search of metagenomic sequencing data against virulence factor database (VFDB) (216) with the cutoff of bitscore 60 was used to calculate the abundance of virulence factors (VFs) in the metagenomic datasets. Both ARG and VF abundances in metagenomes were normalized for metagenomic dataset size and average genome size of the microbial community samples using the frequency of reads encoding the single-copy *rpoB* gene. *rpoB*-encoding reads were identified by a ROCker model, as previously described (217), and the normalized values were presented as genome equivalents (GEs), i.e., fraction of total genomes sampled that encoded the gene of interest (assuming a single-copy for each gene per genome).

For metagenome assembly, paired-end reads were first merged using PEAR (218) (options: -p 0.001) and both merged and unmerged reads were quality-trimmed and human-filtered as described before. These processed merged and unmerged reads from single metagenomes (no co-assembly) were combined for assembly using IDBA-UD (219) (default options). Resulting contigs were binned into metagenome-assembled genomes (MAGs) using MaxBin and their completeness and contamination were assessed with CheckM (220, 221). Quality of MAGs was calculated as "Quality = Completeness – 5 × Contamination", and MAGs with a quality score above 50 were used for further analysis. The Microbial Genome Atlas (MiGA) webserver was used to

determine the most likely taxonomic classification and novelty rank of good-quality MAGs against the classified species in NCBI's prokaryotic genome database (222). To de-replicate the collection of MAGs, we applied a genome-aggregate average nucleotide identity (or ANI) cutoff of 95%, and selected one MAG with the highest quality score as the representative of each resulting 95% ANI-based genomospecies. For the strain level de-replication of MAGs within a genomospecies, a ANI 99% cutoff was used. The relative abundance of MAGs and their encoded ARG in all metagenomes was calculated by competitive bowtie2 alignment (112) and normalized as GEs using ROCker model for *rpoB* as described above. Additionally, for a more conservative estimate, we calculated the 80% truncated coverage for MAG abundance using the BedGraph.tad.rb script of the Enveomics collection (115) to remove outliers genes in terms of coverage such as the rRNA and other multi-copy genes.

4.3.4 Statistical Analyses

β-diversity (i.e., between samples taxa sharing), calculated from Mash distances, was visualized in a non-metric multidimensional scaling (NMDS) plot using the vegan package in R v3.5.1 (223). Permutational multivariate analysis of variance (PERMANOVA) was further used to reveal the effects of measured variables on the Mash distances (adonis function of the vegan package) (223). All differently enriched taxonomic and functional counts between different groups of samples, categorized by each variable, were identified by the Kruskal-Wallis test followed by Dunn's post-hoc test with p-value adjustment with the Benjamini-Hochberg method using the FSA packages (224) in R v3.5.1.

4.3.5 Phylogenetic analysis of MAGs and reference genomes of Escherichia coli

Orthologous genes for *E. coli* MAGs and reference genomes were identified using reciprocal best matches (RBM) using the *rbm.rb* and *ogs.mcl.rb* scripts of the Enveomics collection (115, 225). The core orthologous genes (i.e., present in all genomes) were aligned using Clustal Omega (226) and a phylogenetic tree was built using the concatenated alignment with FastTree version 2.1.7 (227) and default options. The tree was visualized in iTOL (228).

4.4 Results and discussion

4.4.1 Characteristics of the samples and metagenomes

Out of the total 122 individual children sampled, 59 were 1-11 months old, 27 were 12-23 months old and 36 were 24-48 months old at the time of sampling (Table 4-1A). 21 baseline samples were excluded from any comparison that involved sanitation intervention as a variable because all these samples were collected prior to sanitation intervention. Among the remaining 101 samples, 56 samples were exposed to the sanitation intervention (i.e., a sanitation intervention group) while the remaining samples were not (i.e., a control group). We obtained between 5 to 22 million pair-ended reads after trimming per dataset (average read length of 115 to 136 bp per dataset). The estimated abundance-weighted average coverage calculated by Nonpareil for all metagenomes ranged between 81-97%, that is the fraction of the total community DNA that was sequenced, except for one sample that had 91% human contaminating reads. Even this metagenome, however, showed 70% coverage of the sampled microbial community, indicating high enough coverage for assembly and genome binning (Figure

4-1A). Except for this sample, 113/121 metagenomes contained less than 1% of human DNA, six metagenomes contained less than 5% of human DNA, and one contained 25% of human DNA. Assembly and binning effort produced a total of 963 high-quality quality (completeness – 5*contamination ≥ 50) metagenome-assembled genomes (MAGs), representing 153 95%-ANI-based genomospecies (Table C 1). Taxonomic classification through MiGA (222) suggested that 60 genomospecies showed less than 60% of genome-aggregate average amino acid identity (AAI) of the shared genes with any known reference genome, indicating these MAGs represented at least novel genera, if not higher taxonomic ranks (Table C 1). This result corroborates the findings of recent studies that reported distinct gut microbiota in African human populations compared to European populations (208, 229, 230).

Table 4-1. Samples used in this study by group. (A) 59 of the samples were from 1-11 months old children, 27 from 12-23 months and 36 from 24-48 months.

	Time	Group	Treatment	# of samples
Set 1	Baseline	Control (C)	Non-intervention (N)	10
Set 2	Baseline	Intervention (I)	Non-intervention (N)	11
Set 3	Midline	Control (C)	Non-intervention (N)	50
Set 4	Midline	Intervention (I)	Intervention (S)	51

(B) Environmental and health variables measured for each sample.

Metadata	Description

Age	A.1-11months, B.12-23months, C.24-48months
Sanitation intervention	Intervention or Control
Gender	0: male, 1: female
Compound sanitary score	range 0-3 with higher numbers representing riskier conditions (Flooding, leaking wastewater, feces visible around compound grounds).
Measure of household level crowding	binary as 0 for <=3 or 1 for >3
Measure of compound level population density	range 1-5 (higher numbers represent more dense)
Chronic pain	0=no, 1=yes
Diarrhea	0=no, 1=yes

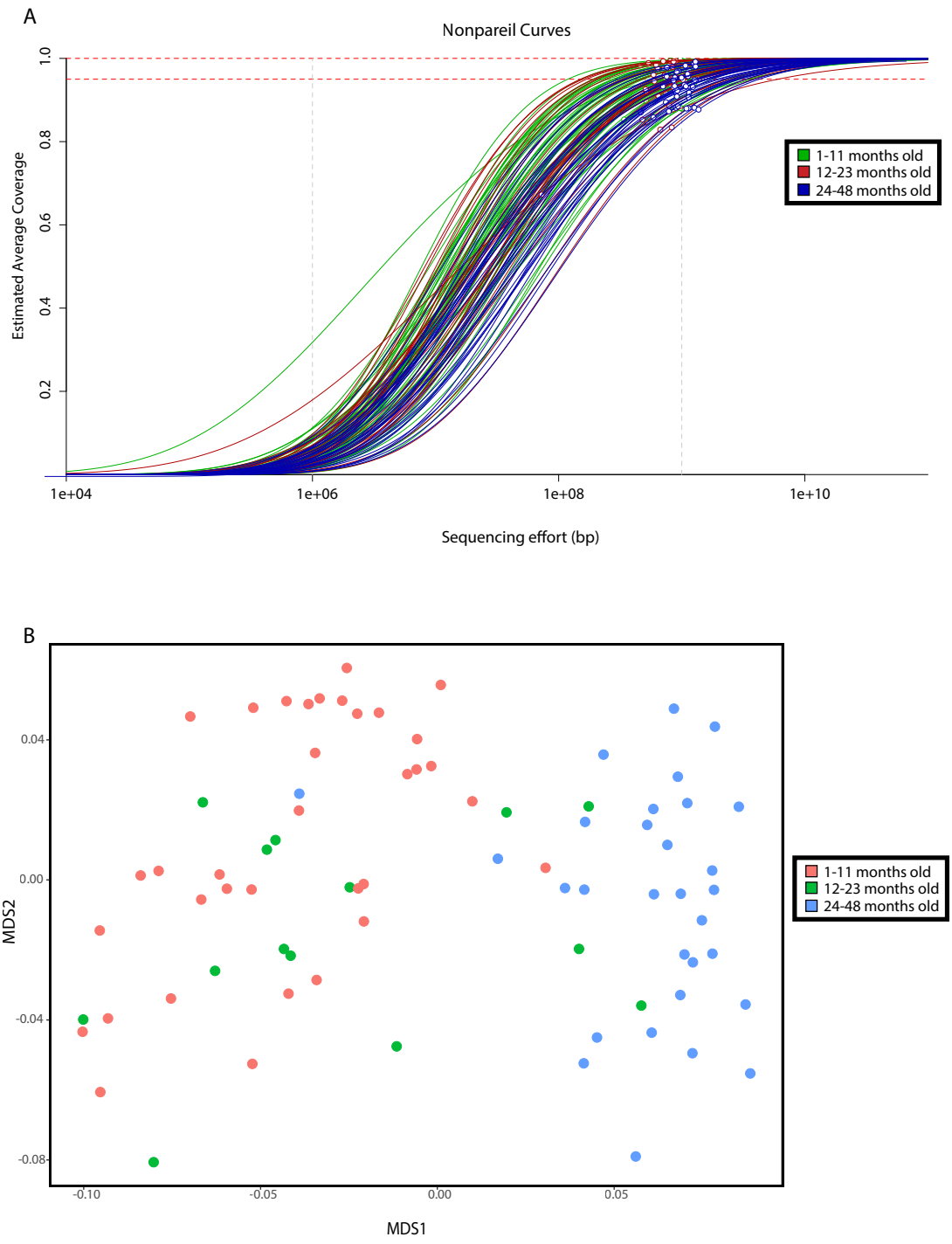


Figure 4-1. Microbial community diversity patterns. (A) Nonpareil curves of the samples used in the study (n=122). The circles of the curves represent the estimated average coverage at the sequencing depth/effort applied. Projected line to the right of the circle represents the expected coverage for higher sequencing efforts. Dashed lines represent the 95%

and 99% coverage levels. Curves positioned more on the right represent more sequence-diverse metagenomes compared to curves positioned on the left.

4.4.2 Microbial community composition

Nonpareil curves (Figure 4-1A), a measure of combined species richness and evenness (total diversity), suggested that microbial communities of the sanitation intervention group show similar diversity to those of the control group (N_d values of 16.88 ± 0.62 vs. 16.81 ± 0.60, respectively; P value of > 0.01, Student's t-test). However, the microbial communities of the 24-48 months group were more diverse than those of the 1-11 months old group (N_d 17.18 ± 0.55 vs. 16.67 ± 0.57, respectively; P value of < 0.001, Student's t-test; note that N_d values are log scale, so a difference of 1 unit translates to 10-fold difference; see also Figure 4-1A). Similarly, Mash, a tool that uses kmer composition for β-diversity calculations, also revealed a clear separation of 24-48 months microbial communities from those of the 1-11 months group. These results are consistent with those of previous studies that reported adult-like microbial composition for children of 3-4 year old but not the communities of younger children (53, 57, 58). No clear difference was observed between the sanitation intervention and the control groups in terms of Nonpareil N_d diversity values (Figure 4-1B). Permutational multivariate analysis of variance (PERMANOVA), a non-parametric test to evaluate the effects of variables through partitioning of variance, was performed in order to investigate the effect of age, sanitation intervention, gender, compound level sanitary conditions, compound level population density, household level crowding, chronic pain, and diarrhea on mash distance of the microbial communities. Interestingly, about 91% of the microbial community variation observed among samples was explained by the set of 8 variables

and their interactions (Table C 2). Specifically, age, diarrhea and a combination of compound sanitary index and diarrhea were significant in explaining diversity patterns (P value of < 0.05, PERMANOVA). Sanitation interventions together with age, compound sanitary index or compound level population density also explained small parts of the variation but not significantly (P value of < 0.1, PERMANOVA). These results revealed that age had the largest effect on the microbial community composition patterns observed between children (R^2 value of 0.315, PERMANOVA) and sanitation intervention also explained small but significant parts of the diversity, especially in combination with the other variables such as age (R^2 value of 0.055, PERMANOVA).

Because the microbial communities of the 1-11 months group showed a clear separation with those of the 24-48 months group, PERMANOVA was also performed for each age group separately in order exclude the effect of age on the Mash distances among the microbial communities. Similar to the results of all age groups reported above, the sanitation intervention on its own explained small parts of the variation but not significantly (P value of < 0.1, PERMANOVA) (Table C 3 and Table C 4), and it explained much larger fraction of diversity in combination with other variables. Interestingly, in the 24-48 months group, other environmental variables such as compound level sanitary conditions, compound level population density, and household level crowding, were significant in explaning variation (P value of < 0.05, PERMANOVA) (Table C 4). These observations suggested that sanitary conditions and population density have significant effects on the microbial community composition patterns especially in the older children who have been under sanitation for longer periods of time and/or are mobile.

4.4.3 Prevalence of virulence factors (VFs) and antibiotic resistance genes (ARGs)

Only the age variable showed significant effect on VF abundance while the effects of the seven other variables evaluated were not significant (Figure 4-2A) (P value of < 0.05, Kruskal-Wallis test). Specifically, both the 1-11 months (total VFs abundance, on median, 0.31; interquartile range [IQR], 0.14 to 0.78 genome equivalents or GEs) and the 12-23 months groups (total VFs, on median, 0.56; IQR, 0.24 to 1.48 GEs) showed significant higher VF abundance compared to that of the 24-48 months group (total VFs, on median, 0.03; IQR, 0.01 to 0.09 GEs) (adjusted P value [P_{adj}] of < 0.05, Dunn's post-hoc test).

Consistent with the VF results, ARG abundance was significantly affected only by age among the eight variables tested (P value of < 0.05, Kruskal-Wallis test), and both the 1-11 months (total ARGs, on median, 4.14; IQR, 3.12 to 6.24 GEs) and the 12-23 months groups (total ARGs, on median, 6.47; IQR, 3.57 to 13.05 GEs) showed significantly higher ARG abundance compared to that of the 24-48 months group (total ARGs, on median, 2.58; IQR 2.56 to 3.23 GEs) (P_{adj} of < 0.05, Dunn's post-hoc test) (Figure 4-2B). Various classes of ARGs were detected in all metagenomes, with multidrug, tetracycline, β-lactam and bacitracin showing the highest abundances in most of the metagenomes (Figure C 1). Because the Mozambican Ministry of Health has recommended the use of ampicillin and gentamicin (i.e., aminoglycoside class) for children under the age of 2, and chloramphenicol for older children (231), we further attempted to investigate the abundances of specific ARGs representative of these classes of antibiotics. Similar to the results mentioned above, only age affected the abundance of multidrug and aminoglycoside resistance genes (P value of < 0.05, Kruskal-Wallis test),

and higher abundances were observed in the 1-11 months (aminoglycoside and multidrug resistance gene abundance, on median, 0.16; IQR, 0.09 to 0.5 GEs and 1.24; IQR, 0.58 to 2.33 GEs, respectively) and the 12-23 months group (aminoglycoside and multidrug resistance gene abundance, on median, 0.45; IQR, 0.17 to 1.04 GEs and 2.10; IQR, 0.84 to 6.27 GEs, respectively) compared to the 24-48 months group (aminoglycoside and multidrug resistance gene abundance, on median, 0.03; IQR, 0.02 to 0.06 and 0.14; IQR 0.06 to 0.56 GEs, respectively) (P_{adj} of < 0.05, Dunn test) (Figure C 2A). This result is consistent with the (recommended) usage of gentamicin for children <2 year old. In case of β-lactam class, only the gender variable affected significantly the relative abundance of the corresponding ARGs (P value of < 0.05, Kruskal-Wallis test), while the age was marginally significant (P value of ~ 0.078, Kruskal-Wallis test) (Figure C 2B). Sanitation intervention was the only variable that showed significance on the relative abundance of chloramphenicol resistance genes (P value of < 0.05, Kruskal-Wallis test). Specifically, the control group showed lower abundance of chloramphenicol resistance genes (median of 0.005; IQR, 0.002 to 0.013 GEs) compared to the sanitation intervention group (median of 0.012; IQR, 0.003 to 0.050 GEs) (Figure C 2C). Additionally, taxonomic affiliation of the bacterial genomes that encoded the ARG genes was also investigated. 74 to 99% of the metagenomic reads encoding ARG were assignable to a set of MAGs representing 153 genomospecies (Figure C 3 and Table C 1). Only two genomospecies (i.e., ANIsp_33: best AAI match to *Alistipes_finegoldii* with 66.79% and ANIsp_46: best AAI match to *Faecalibacterium_prausnitzii* with 52.42%) showed significant difference between the control and sanitation intervention groups (P value of < 0.05, Kruskal-Wallis test); these genomospecies (and their encoded ARGs) showed slightly higher abundance

in the sanitation intervention (total ARG abundance, on median, 0; IQR, 0 to 0.004 and 0; IQR, 0 to 0.009 GEs in ANIsp_33 and ANIsp_46, respectively) vs. the control groups (total ARG abundance, on median, 0; IQR 0 to 0 and 0; IQR 0 to 0.0004 GEs in ANIsp_33 and ANIsp_46, respectively) (**Figure C 4**). Overall, these results suggested that sanitation intervention might not have a significant effect on the relative abundance of VFs or ARGs in the gut microbiome of Mozambican children.

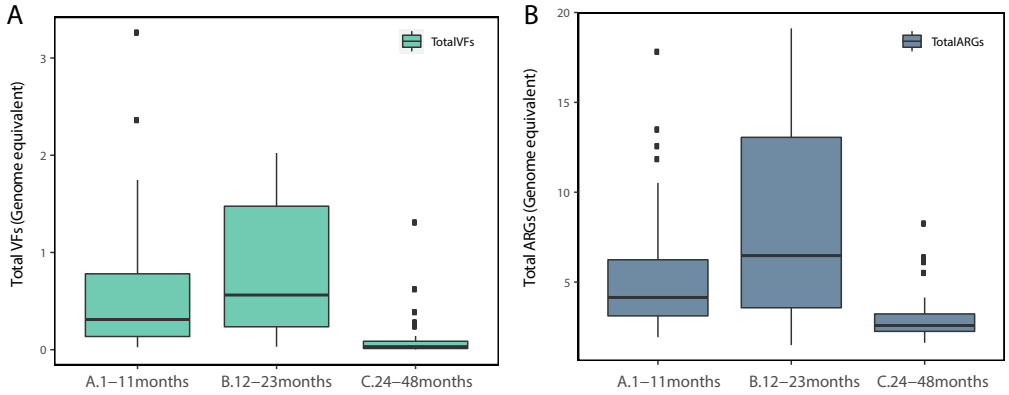


Figure 4-2. Relative abundance of VFs and ARGs in multiple age groups. (A) Total VFs abundance in three different age groups. (B) Total ARGs abundance in three different age groups. Abundance (y-axes) was defined as genome equivalents, i.e., what fraction of total cells/genomes sampled encoded the gene of interest, assuming that the genes are single-copy in the genome. The line that divides the box into 2 parts represents the median of the data. The end of the box shows the upper and lower quartiles. The extreme line shows the highest and lowest value, excluding outliers.

4.4.4 Effect of sanitation intervention on species abundance

We examined species abundance differences between the sanitation intervention and the control groups to test the effect of the sanitation intervention on specific gut taxa and enteric pathogens using two distinct approaches, MetaPhlAn2 species profiles and read-recruitment analysis with the recovered MAGs, including MAGs representing

unknown (not previously described) genomospecies. All good quality MAGs, representing 153 genomospecies, collectively recruited between 40 to 97% of the total reads, depending on the metagenome considered (Table C 1and Figure C 5). Out of 153 genomospecies, seven genomospecies showed significant difference in abundance between the control and the sanitation intervention groups (P value of < 0.05, Kruskal-Wallis test) (Figure 4-3A). ANIsp_39 (closely related to *Bacteroides_thetaiotaomicron* with 98.13% AAI), ANIsp_46 (distantly related to *Faecalibacterium_prausnitzii* with 52.42 % AAI), and ANIsp_84 (moderately related to *Blautia_coccoides* with 68.02% AAI) were more abundant in the sanitation intervention vs. the control groups (median coverage of 0.00001; IQR, 0 to 0.003 vs. 0; IQR, 0 to 0.00045, 0; IQR, 0 to 0.0001 vs. 0; IQR 0 to 0 ± 0.0004, and 0; IQR 0 to 0 vs. 0; IQR 0 to 0 GEs, respectively) (Figure 4-3). On the other hands, ANIsp_93 (distantly related to *Paeniclostridium_sordellii* with 40.50% AAI), ANIsp_100 (distantly related to *Ruminiclostridium*_sp with 48.62% AAI), ANIsp_130 (closely related to *Veillonella_parvula* with 95.27% AAI), and ANIsp_143 (distantly related to *Streptococcus_parasanguinis* with 46.08% AAI) were less abundant in the sanitation intervention vs. the control groups (median coverage of 0; IQR 0 to 0.00015 vs. 0.0002; IQR, 0 to 0.0018, 0; IQR 0 to 0 vs. 0; IQR 0 to 0.000008, 0; IQR 0 to 0 vs. 0; IQR, 0 to 0.0005, and 0; IQR 0 to 0 vs. 0; IQR, 0 to 0.00015 GEs, respectively) (Figure 4-3).

ANIsp_039, whose closest relative was *B. thetaiotaomicron* type strain VPI-5482 with 98.13% of AAI, showed higher abundance in the sanitation intervention compared to the control group as mentioned the above. In addition to this, two other variables such as chronic pain and household level crowding also significantly affected that abundance

of *B. thetaiotaomicron* (P value of < 0.05, Kruskal-Wallis test) (Figure 4-3B), while age, gender, compound level sanitary conditions and compound level population density were not strongly effecting the abundance of this species (P value of > 0.05, Kruskal-Wallis test). More specifically, *B. thetaiotaomicron* showed higher abundance in the group having equal or less than 3 people in a household vs. the group having more than 3 people in a household (median coverage of 0.000045; IQR, 0 to 0.002 vs. 0; IQR 0 to 0 GEs, respectively), and also showed higher abundance in the presence of chronic pain vs. the absence of chronic pain groups (median coverage of 0; IQR, 0 to 0.0009 vs. 0.006; IQR 0.005 to 0.008 GEs, respectively) (Figure 4-3B). However, it should be noted that the number of samples in the chronic pain group is relatively low (n = 5). Hence, the results about chronic pain should be considered only as preliminary. *B. thetaiotaomicron* type strain VPI-5482 was originally isolated from the feces of a healthy adult human (232), and it has been reported to stimulate angiogenesis during postnatal intestine development resulting in increased capacity of the (human) host to absorb nutrients in germfree mice (233). Moreover, *B. thetaiotaomicron* regulates the synthesis of various gut epithelial glycans (234). However, this normally beneficial gut microbial species has also been shown to increase the virulence and disease progression of enteric pathogens such as enterohemorrhagic *Escherichia coli* (EHEC) by altering the metabolic landscape within the gut during murine infection, when the two species co-occur at a frequency of 10:1 (235). Overall, it appeared that the combination of sanitation interventions and less household level crowding might have an effect to the (increased) abundance of *B. thetaiotaomicron* in the gut microbiota, which is generally beneficial to the host.

V. parvula can be found in normal flora of the mouth, gastrointestinal tract, and vagina in humans but has also been implicated in various infections of the nose, lungs, heart, bone and central nervous system (236), and has also been isolated from bacterial vaginosis samples (237). Interestingly, *V. parvula* (ANIsp_130, Table C 1) showed higher abundance in the control vs. sanitation intervention groups (median coverage of 0; IQR, 0 to 0.0005 vs. 0; IQR, 0 to 0 GEs, respectively) and the diarrheal vs. the non-diarrheal groups (median coverage of 0.0002; IQR, 0 to 0.001 vs. 0; IQR, 0 to 0.00003 GEs, respectively) (Figure 4-3), while the other 6 assessed variables (age, gender, compound level sanitary conditions, compound level population density, household level crowding, and chronic pain) did not affect the abundance of *V. parvula* (P value of > 0.1, Kruskal-Wallis test). Consistently, MetaPhlAn2 results showed that the abundance of *V. parvula* was not only higher in the control group compared to the sanitation intervention group (median percentage of relative abundance of 0.063; IQR, 0 to 0.18 vs. 0.012; IQR, 0 to 0.05, respectively) but also was higher in the diarrhea group compared to the non-diarrhea group (median percentage of relative abundance of 0.12; IQR, 0.04 to 0.59 vs. 0.016; IQR, 0 to 0.07, respectively) (P value of < 0.05, Kruskal-Wallis test) (Figure 4-4). These results indicated that *V. parvula* might be an indicator of a non-healthy or abnormal intestinal microbiota in Mozambican children that its abundance is significantly affected by the sanitation interventions. Furthermore, similar to *V. parvula*, two additional novel taxa, ANIsp_93 (distantly related to *Paeniclostridium sordellii* with 40.50% AAI) and ANIsp_143 (distantly related to *Streptococcus parasanguinis* with 46.08% AAI), were more abundant in the control and the diarrhea group than their counterparts (Figure 4-3). These observations suggested that these previous three taxa can

be the potential biomarkers for the effect of sanitation intervention on health condition. However, further experimental work is needed in order to corroborate these findings and elucidate the exact role of these taxa in the gut microbiome of Mozambican children.

MetaPhlAn2 analysis identified a total of 397 microbial species including pathogens such as *Camplybacter jejuni*, *Shigella flexneri*, *Shigella sonnei*, *Salmonella enterica*, *Clostridium difficile* and Human adenovirus, together with commensal taxa. Out of total 397 species, the relative abundance of 10 species (i.e., *Anaerotruncus colihominis*, *Bifidobacterium angulatum*, *Clostridium symbiosum*, *Eubacterium limosum*, Specifically, *Lachnospiraceae bacterium*, *Lactobacillus pentosus*, *Mitsuokella multacida*, *Escherichia sp.* (unclassified), *Lactobacillus ruminis* and *Veillonella parvula*) was different between the sanitation intervention group and the control group (P value of < 0.05, Kruskal-Wallis test). *A. colihominis*, *C. symbiosum*, *Escherichia sp.*, *E. limosum*, and *L. bacterium* were more abundant in the sanitation intervention vs. the control groups (median percentage of relative abundance of 0; IQR 0 to 0 vs. 0; IQR 0 to 0, 0; IQR 0 to 0.015 vs. 0; IQR 0 to 0, 0.46; IQR, 0.09 to 1.61 vs. 0.04; IQR, 0 to 0.84, 0; IQR, 0 to 0 vs. 0; IQR, 0 to 0, and 0; IQR, 0 to 0 vs. 0; IQR, 0 to 0, respectively). On the other hands, *B. angulatum*, *L. pentosus*, *L. ruminis*, *M. multacida*, and *V. parvula* were less abundant in the sanitation intervention vs. the control groups (median percentage of relative abundance of 0; IQR, 0 to 0 vs. 0; IQR, 0 to 0, 0; IQR, 0 to 0 vs. 0; IQR, 0 to 0, 0; IQR, 0 to 0 vs. 0; IQR, 0 to 0.91, 0; IQR, 0 to 0 vs. 0; IQR, 0 to 0, and 0.012; IQR, 0 to 0.05 vs. 0.06; IQR, 0 to 0.18, respectively).

Especially, *Escherichia sp.*, *V. parvula* and *L. ruminis* appeared to be highly enriched in both the sanitation intervention and the control groups, i.e., >0.1% of the total

metagenomes. *L. ruminis,* which was from the feces of humans and other animals (238, 239) and is considered to be a member of the autochthonous gut microbiota with possible positive roles in host-microbe crosstalk (240-242), was less abundant in the sanitation intervention relative to the control group (Figure 4-4A). Notably, we were also able to recover MAGs related to the *L. ruminis* (ANIsp_44 with 96.57% AAI, Table C 1), but its abundance was not significant different between the control and the sanitation intervention groups, which was not consistent with the results of MetaPhlAn2 analysis. Thus, the relevance of *L. ruminis* as a biomarker of the effect of sanitation interventions on gut microbiome remained unclear.

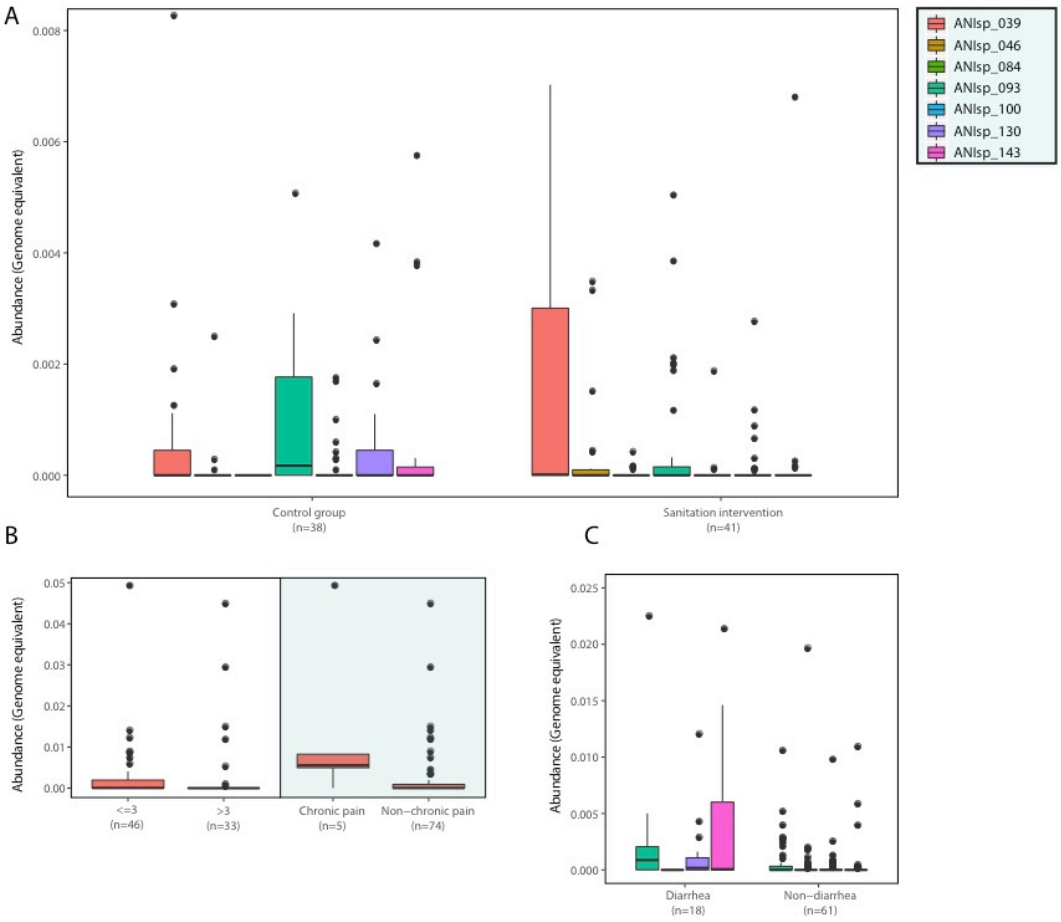


Figure 4-3. Differently abundant species between the control and the sanitation intervention groups (MAG). Abundance (y-axes) was defined as genome equivalents, i.e., how many of the total cells/genomes samples represent a certain MAG ,assuming that the coverage of *rpoB* (a single copy gene) represents the size/coverage of whole microbial community. (A) Average relative abundance of 7 genomospecies that were differentially abundant between the control and sanitation intervention groups are shown, i.e., ANIsp_39 (closest relative to *Bacteroides_thetaiotaomicron* with 98.13% AAI), ANIsp_46 (closest relative to *Faecalibacterium_prausnitzii* with 52.42 % AAI), ANIsp_84 (closest relative to *Blautia_coccoides* with 68.02% AAI), ANIsp_93 (closest relative to *Paeniclostridium_sordellii* with 40.50% AAI), ANIsp_100 (closest relative to *Ruminiclostridium*_sp with 48.62% AAI), ANIsp_130 (closest relative to *Veillonella_parvula* with 95.27% AAI), and ANIsp_143 (closest relative to *Streptococcus_parasanguinis* with 46.08% AAI). Values above 0.008 were not shown here for clarity but included in all statistical tests. (B) Boxplots of the relative abundance of ANIsp_039 (closest relative to *Bacteroides_thetaiotaomicron* with 98.13% AAI) in two

household level crowded groups and two diarrheal groups. (C) The abundance differences of ANIsp_093 (closest relative to *Paeniclostridium_sordellii* with 40.50% AAI), ANIsp_100 (closest relative to *Ruminiclostridium*_sp with 48.62% AAI), ANIsp_130 (closest relative to *Veillonella_parvula* with 95.27% AAI) and ANIsp_143 (closest relative to *Streptococcus_parasanguinis* with 46.08% AAI) between the diarrhea group and the non-diarrhea group are shown.

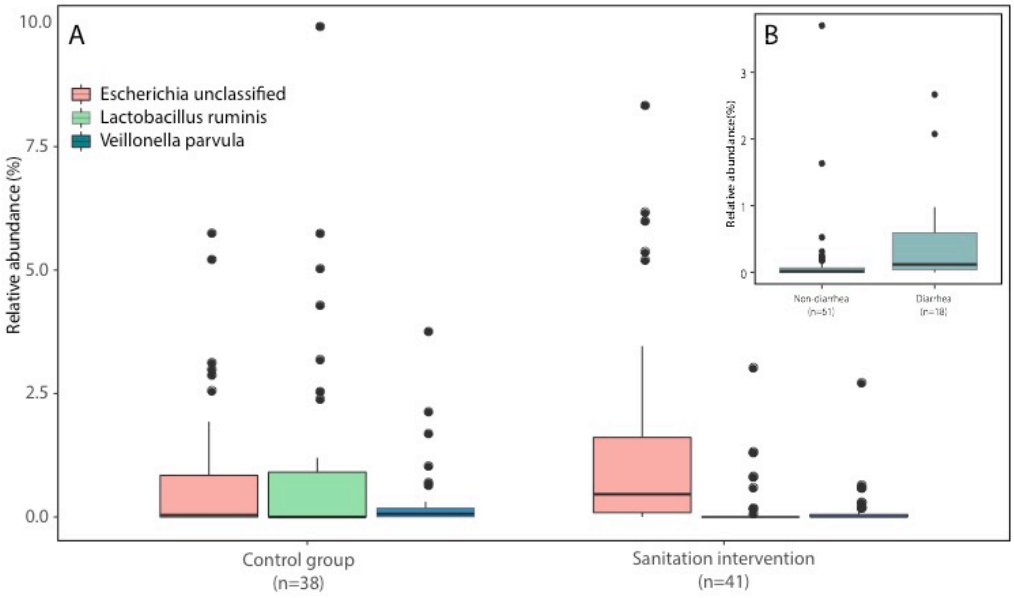


Figure 4-4. Differently abundant species between the control and the sanitation intervention groups (MetaPhlAn2). (A) Boxplots of the relative abundance of *Escherichia sp.*, *Lactobacillus ruminis* and *Veillonella parvula* in the control group and sanitation intervention groups are shown. (B) A boxplot of the relative abundance of *V. parvula* between the diarrhea group and non-diarrhea group (inset) is shown.

4.4.5 Escherichia coli in the gut of Mozambican children

Typically, *E. coli* is an important member of the normal intestinal microflora of humans and other mammals and an early colonizer of infant intestines (243, 244), but it can also be a highly versatile or even deadly pathogen by acquiring specific virulence attributes (245). For example, The Global Enteric Multicenter Study (GEMS), which

identified the etiology and population-based burden of pediatric diarrheal disease in sub-Saharan Africa and South Asia, reported that enterotoxigenic *E. coli* (ETEC) and typical enteropathogenic *E. coli* (EPEC) are associated with increased risk of death in infants aged 0-11 months (196). Additionally, a recent study suggested that enteroaggregative *E. coli* (EAEC) may be the leading cause of diarrheagenic *E. coli* associated food and water-borne enteric infection in South Africa by assessing 205 *E. coli* strains isolated from producer distributor bulk milk, irrigation water, irrigated lettuce and street vendor coleslaw (246).

A total of 70 good-quality MAGs were identified as *E. coli* (>95% ANI to known *E. coli* genomes) and represented 14 distinct strains (<99% ANI among the strains). The abundance of these MAGs was higher in the 1-11 months (median of 0.02; IQR, 0.005 to 0.053 GEs) and 12-23 months (median of 0.05; IQR, 0.005 to 0.17) groups compared to the 24-48 months group (median of 0.00002; IQR, 0 to 0.003 GEs) (P_{adj} of < 0.05, Dunn's post-hoc test) (**Figure 4-5**B) and other measured variables than age did not significantly explain this difference in abundance (P value of > 0.05, Kruskal-Wallis test). Individual *E. coli* strain showed, in general, similar abundance patterns to the total *E. coli* (data not shown here); thus, the abundance of each of the 14 strains of *E. coli* was also not affected by sanitation intervention (P value of > 0.05, Kruskal-Wallis test). Phylogenetic placement of all orthologous genes for the 14 strains and reference *E. coli* strains revealed several clusters (**Figure 4-5**A). Interestingly, four strains clustered together with *E. coli* O44:H18 strain 042, an EAEC epidemic strain (**Figure 4-5**A). Except for one strain (Map4bin002), which was recovered from a baseline intervention sample (i.e., prior to the exposure of sanitation intervention), three other strains (i.e.,

Map48bin004, Map50bin014, and Map49bin008) were recovered from the sanitation intervention group. However, all four strains (i.e., Map4bin002, Map48bin004, Map50bin014, and Map49bin008) showed, more or less, similar abundances between the sanitation intervention and the control groups, consistent with the total *E. coli* populations data (P value of > 0.05, Kruskal-Wallis test), as mentioned the above.

In order to investigate more closely the pathogenic potential of the recovered *E. coli* MAGs, we collected several diagnostic virulence genes of *E. coli* (Table C 5) and compare their presence/absence and relative abundance between the control and the sanitation intervention groups. 48 samples out of 79 samples showed a positive signal of ETEC and this was the most prevalent among the five known pathotypes included in our analysis. Only two variables, age and diarrhea, were strongly influenced the abundance of the ETEC diagnostic genes (P value of < 0.05, Kruskal-Wallis test). Interestingly, we found that the diarrhea group (median of 0.003; IQR, 0.0009 to 0.006 GEs) showed a higher abundance of ETEC diagnostic genes compared to the non-diarrhea group (median of 0.00005; IQR, 0 to 0.003 GEs) (P value of < 0.05, Kruskal-Wallis test) (**Figure 4-5**C). This result is in agreement with previous studies that ETEC is one of the important causative agents of diarrhea-associated morbidity and mortality among children of age five or younger (196, 247). However, sanitation intervention was not strongly associated with the prevalence of ETEC in Mozambican children. The abundance of the diagnostic genefor typical EPEC (i.e., *bfpA*, Table C 5) was associated with the sanitation intervention and were, in fact, higher in the sanitation intervention group (median of 0; IQR, 0 to 0 GEs) compared to the control group (median of 0; IQR, 0 to 0 GEs) (P value of < 0.05, Kruskal-Wallis test), while the abundance of other diagnostic genes were not

significantly associated with sanitation intervention (P value of > 0.05, Kruskal-Wallis test). It is important to note, however, that only five of samples in the intervention group showed a positive signal for typical EPEC, while the remaining 74 samples showed zero abundance, which indicated that the results obtained may be heavily influenced by a relative small number of individuals sampled. In the case of EAEC, the same MAGs (at 99% ANI level) was recovered from several samples in the sanitation intervention group, and the abundance of EAEC diagnostic genes was slightly higher in the sanitation intervention group (median of 0; IQR, 0 to 0.00006 GEs) compared to the control group (median of 0; IQR, 0 to 0 GEs) but this difference was not statistically significant (P value of ~ 0.07, Kruskal-Wallis test). Additionally, the abundance of all diagnostic genes was not strongly correlated with diarrhea except for the ETEC diagnostic genes (P value of > 0.05, Kruskal-Wallis test). These results suggested that the sanitation intervention did not significantly affect the abundance of the major *E. coli* pathotype groups (i.e., ETEC, typical and atypical EPEC, EAEC, DAEC, and EHEC) and that *E. coli* relative abundance was relative high in most children sampled.

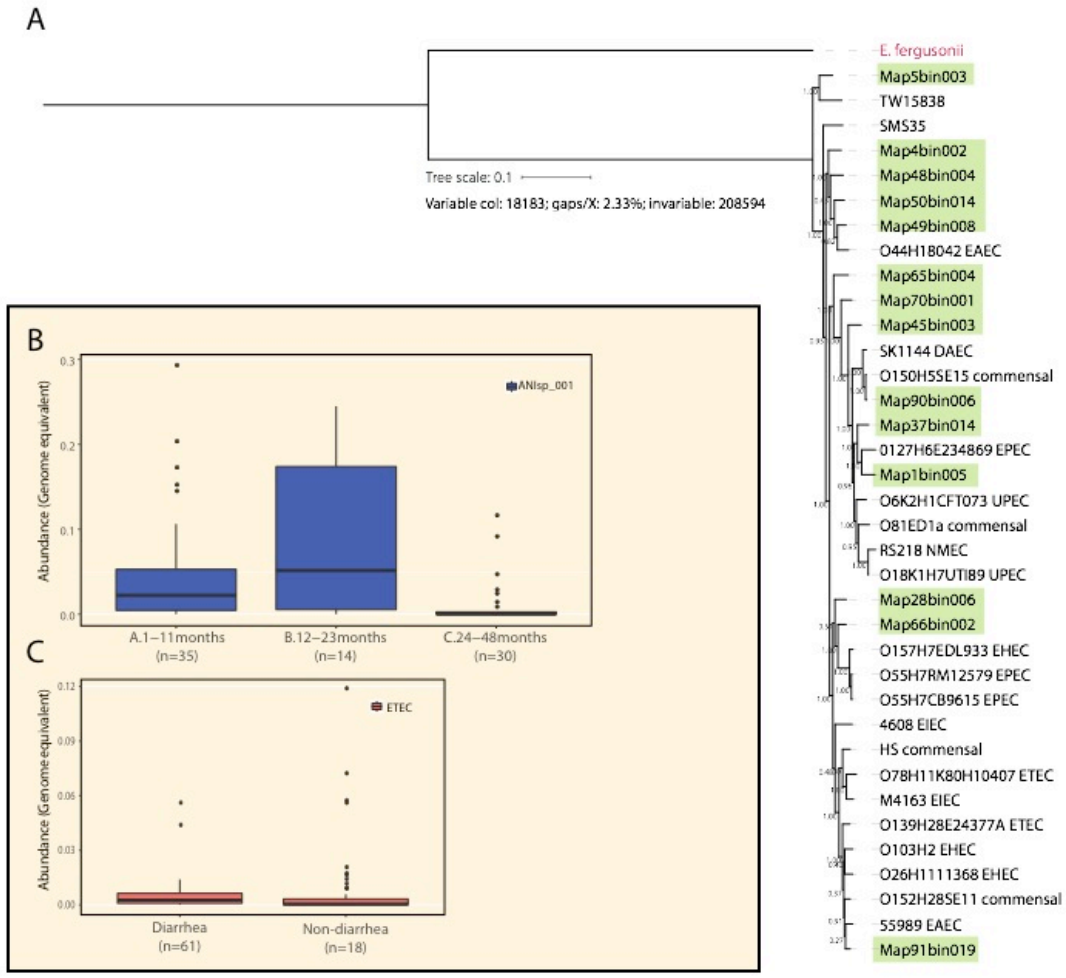


Figure 4-5. *Escherichia coli* MAGs in gut microbiota of Mozambican children. (A) Phylogenetic tree of *E. coli* orthologous genes. The unus genome-based, which is the collection of orthologus groups present in single copy in all genomes, phylogeny of *E. coli* MAGs and selected reference genomes are shown, including commensal (**HS commensal**, CP000802.1 strain HS; **O150H5SE15 commensal**, AP009378.1 strain SE15; **O152H28SE11 commensal**, AP009240.1 strains SE11; **O81ED1a commensal**, CU928162.2 strain ED1a), pathogenic (**0127H6E234869 EPEC**, FM180568.1 strain 0127:H6 E2348/69; **4608 EIEC**, gi|735003713|gb|JTCO01000001.1| strain 4608-58 4608-58_c1; **55989 EAEC**, gi|218350208|emb|CU928145.2| strain 55989; **M4163 EIEC**, gi|735003727|gb|JTCN01000001.1| strain M4163 M4163_c1; **O103H2 EHEC**, AP010958.1 strain O103:H2 str. 12009; **O139H28E24377A ETEC**, CP000800.1 strain E24377A; **O157H7EDL933 EHEC**, gi|749302083|ref|NZ_CP008957.1| strain O157:H7 str. EDL933; **O18K1H7UTI89 UPEC**, CP000243.1 strain UTI89; **O26H1111368 EHEC**, AP010953.1 strain O26:H11 str. 11368; **O44H18042 EAEC**, gi|284919779|emb|FN554766.1| strain 042; **O55H7CB9615 EPEC**, CP001846.1 strain O55:H7 CB9615; **O55H7RM12579 EPEC**, CP003109.1 strain O55:H7 RM12579; **O6K2H1CFT073 UPEC**, AE014075.1 strain CFT073; **O78H11K80H10407 ETEC**,

FN649414.1 strain ETEC H10407; **RS218 NMEC**, CP007149.1 strain RS218; **SK1144 DAEC**, NZ_AP018784.1 strain SK1144), and environmental strains (**SMS35**, gi|170517292|gb|CP000970.1| strain SMS-3-5; **TW15838**, gi|329753645|gb|AEJX01000001.1| *E. sp*. TW15838). (B) The abundance of *E. coli* MAGs in multiple age groups. (C) Higher abundance of ETEC diagnostic genes in the diarrhea group compared to the non-diarrhea group.

4.5 Conclusions and outlook

In this study, we applied a shotgun sequencing approach and associated bioinformatic pipelines to examine the effect of several environmental and health-related variables on the gut microbiome composition of children under 48 months old in a low-income, urban neighborhood in Maputo, Mozambique. Our major findings suggested that the dynamics of gut microbial community diversities (i.e., both α- and β-diversity), ARG abundance and VF abundance were mostly influenced by age group, and secondarily by sanitation or diarrheal status. Consistently, we were able to identify only a few species and gene functions that were differentially abundant between the sanitation intervention and the control group. Most notably, the higher abundance of *V. parvula* in the control vs. the sanitation intervention groups and in the diarrhea vs. the non-diarrhea groups, and the higher abundance of a few commensal taxa such as *B. thetaiotaomicron* in the sanitation intervention group were observed. Considering that *V. parvula* has been implicated as a pathogen in various infections including bacterial vaginosis in human, this species might be a potential biomarker for assessing health in Mozambican children.

It should be noted that there were several limitations in our study that might have affected our results and conclusions. Most notably, all participating households that had access to the shared sanitation facility without safe excreta management, which is a key criterion used in assessing eligibility for the MapSan trial (48). In such settings, we expected to see high exposure from various sources (soil, water, person-to-person

contact) that can mask the effect of sanitation intervention. Moreover, in this set of samples, only 1-11 months old children in the sanitation intervention group were exposed to the improved sanitation facilities for their entire lifetime, while the other two older-age groups, i.e., 12-23 months and 24-48 months, had been exposed to the poor sanitation at the early stage of their life (followed by sanitation intervention that last ~11 months). It is thus possible that these older children had been colonized by certain pathogens during their early life (when they did not experience sanitation intervention), resulting in potential health problems and long-term pathogen colonization Furthemore, the number of samples analyzed here is still limited; future efforts should try to expand on the children sampled and the duration that the children experience sanitation interventions (196).

Overall, it appears that the sanitation intervention employed in this study might have a significant effect on reducing health risks in Mozambican children but the effect is still rather minor and maybe longer sampling times or proper sets of samples, e.g., children exposed to sanitation interventions for their whole life time, are needed to more robustly quantify the effect. Furthermore, the commensal taxa recovered here might have important (unknown) roles in the gut microbiome functioning related to sanitation intervention, which should be subject of future research.

4.6 Acknowledgements

This work was supported by the US NIH (award 1R01AI130163-01A1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

APPENDIX A SUPPLEMENTARY MATERIAL FOR CHAPTER 2

A.1 Supporting results and discussion

A.1.1 Integrative and conjugative elements (ICEs)

In contrast to plasmids, ICEs can be integrated into the host chromosome via phage integrase-mediated chromosomal integration at specific sites encoding tRNA genes (248). In GI-2, we found a tyrosine-based site-specific recombinase CMGI-2, and a $tRNA^{Gly}$, which was located 80 bp upstream of GI-2 and overlapped, over 22bp, with a sequence that was duplicated at the other end of GI-2 (**Figure A 3**). These results suggested that tyrosine-based recombinase might have catalyzed the site-specific integration of GI-2 in *P. aeruginosa* DPB. However, GI-2 also harbored a transposase DDE domain protein, which is known to employ various mechanisms of recombination with promiscuous integration sites (249), indicating that non-site-specific integration is also possible for this ICE. Furthermore, the presence of plasmid replication initiator (*repA*) in GI-2 suggested that GI-2 may be also prone to plasmid-like replication, similar to several other known ICEs (250), which blur the line dividing ICEs and conjugative plasmids.

A.1.2 BAC-adaptation in P. aeruginosa

In contrast to all *P. aeruginosa* DP BAC(+) populations, *P. aeruginosa* DPB BAC(+) and BAC(-) populations did not showed higher MIC to the membrane-active antibiotic polymyxin B except for DPB BAC(+)_1 compared to the ancestor even thought they all had mutations in *pmrB*, albeit at different locations than DP BAC(+)

(Table 2-2 and **Table A 3**B). The mutations to *pmrB* in DPB BAC(+) and BAC(-) populations apparently did not affect their polymyxin B resistance level relative to their ancestor (but only their BAC tolerance level) because the level of polymyxin resistance depends on the type of mutation in *pmrB* (e.g., exact location and combinations of mutation) and possibility of presence of other determinants for resistance phenotype (137). For instance, repeated passages without polymyxin in the growth media results in loss of resistance in some cases, suggesting that the *pmrAB* locus is not the only determinant of the resistance phenotype (137). These other (unknown) determinants and their epistatic effects on the mutation on the *pmrB* gene presumably accounted for the lack of increased polymyxin B resistance in DPB BAC(+) populations. Consistent with these interpretations, we found that the expression level of *arnBCADTEF* operon under BAC-free (but not under BAC-exposed) conditions in all DPB BAC(+) populations was not significantly different from that of their ancestor, and almost no expression of *arnE* and *arnF* (Table A 6A). However, when BAC was added to the growth media before taking the sample for transcriptomics, overexpression of *arnBCADT* and almost no expression of *arnE* and *arnF* were observed (Table A 6B). *arnE* and *arnF* are presumably responsible for transportation of undecaprenyl phosphate-α-L-Ara4N from inner membrane to outer surface (138), which is required for the last step of the synthesis of L-Ara4N modified lipidA. Therefore, it appears that, at least in *P. aeruginosa* DPB, overexpression of *arnBCAD* produces undecaprenyl phosphate-α-L-Ara4N in the inner membrane, which is important for increased BAC tolerance but not relevant for conferring polymyxin B resistance; the latter apparently requires the transportation of the undecaprenyl phosphate-α-L-Ara4N to the outer membrane. And, incomplete expression

of *arnBCADTEF* operon, especially *arnE* and *arnF*, presumably accounted for the fact that DPB BAC(+) populations did not show increased polymyxin B resistance under BAC exposed condition compared to their ancestor (**Table A 3**C). Consistent with these findings and interpretations, the DPB BAC(-)_1 population, which showed decreased BAC tolerance relative to the other two *P. aeurginosa* DPB BAC(-) replicate populations, had 20bp insertion mutation in *pmrA*, and consequently the expression level of *arnBCADTEF* operon was not different from that of ancestor or control under the BAC-exposed condition (Table A 4D and Table A 6C).

All *P. aeruginosa* DP BAC(-) populations showed loss of polymyxin B resistance relative to DP BAC(+) populations even though they had fixed mutations in *pmrB* (Figure 2-3A and Table 2-2). Read alignment of DP BAC(-) datasets against assembled contigs of *P. aeruginosa* DP ancestor suggested that subpopulations of *P. aeruginosa* DP BAC(-)_2 and DP BAC(-)_3 had different mutations in *pmrB* and *pmrA* compared to those in *P. aeruginosa* DP BAC(+) populations (**Table A 5**). This finding may explain the loss of polymyxin B resistance in these populations since the location and combination of mutations in *pmrB* are known to affect the level of polymyxin B resistance (discussed above). In the case of population *P. aeruginosa* DP BAC(-)_1, which did not have the mutations in *pmrB* or *pmrA* that P. aeruginosa DP BAC(-)_2 and BAC(-)_3 had, it is likely that mutations elsewhere in the genome during the BAC(-) phase were responsible for loss of polymyxin B resistance. For instance, this population uniquely had a 15bp insertion mutation in the quorum sensing regulator gene *lasR* (Table A 4C). Expression of more than 300 genes, including MFS transporters, RND efflux pump genes, and two-component response regulators, are regulated by two acyl-

homoserine lactone systems, LasR-LasI and RhlR-RhlI (166, 167). Therefore, it is likely that the mutation in *lasR* might have induced expression changes in the genes related to polymyxin B resistance in *P. aeruginosa* BAC(-)_1 population, although this hypothesis awaits experimental validation. In summary, BAC exposure selected for the mutations in *pmrB* that conferred increased BAC-tolerance in all *P. aeruginosa* populations, but only specific mutations conferred increased polymyxin resistance in the DP (but not the DPB) genetic background. For the remaining of mutations, gene expression data provided testable hypotheses for explaining the lack of increased polymyxin B resistance.

A.2 Supplementary figures and tables

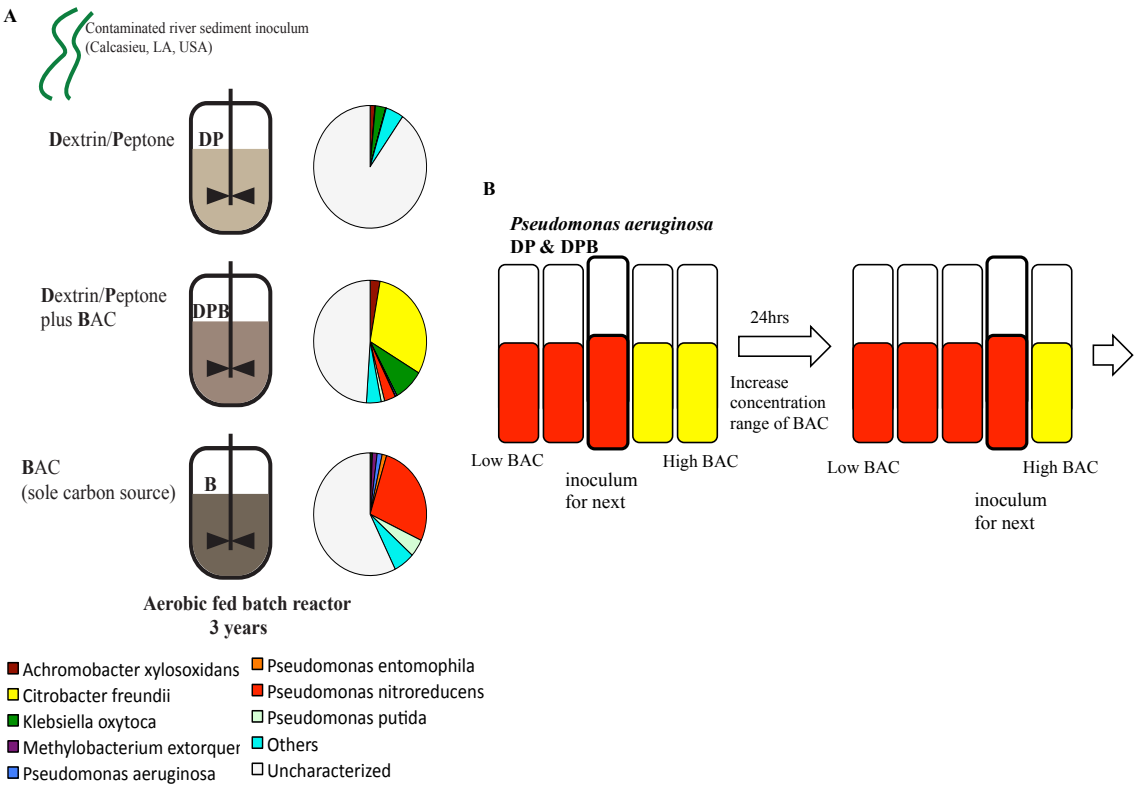
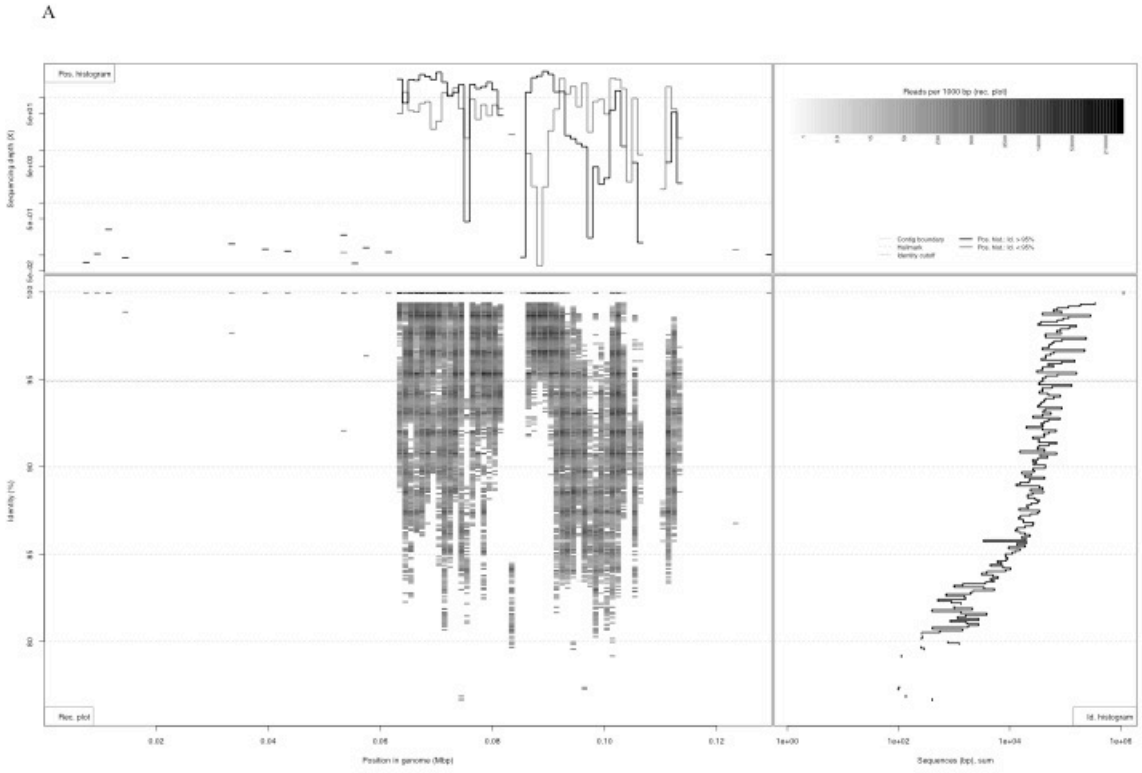
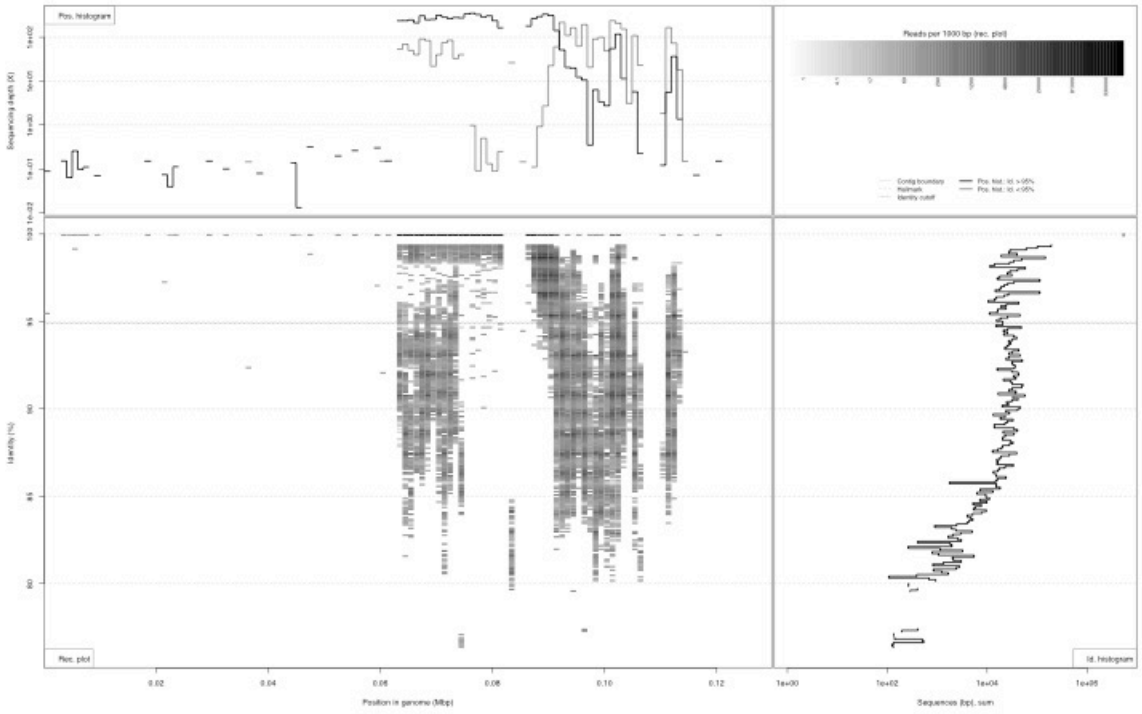


Figure A 1. Bioreactor development and adaptive evolution experiment. (A) Bioreactors were built with a microbial community inoculum

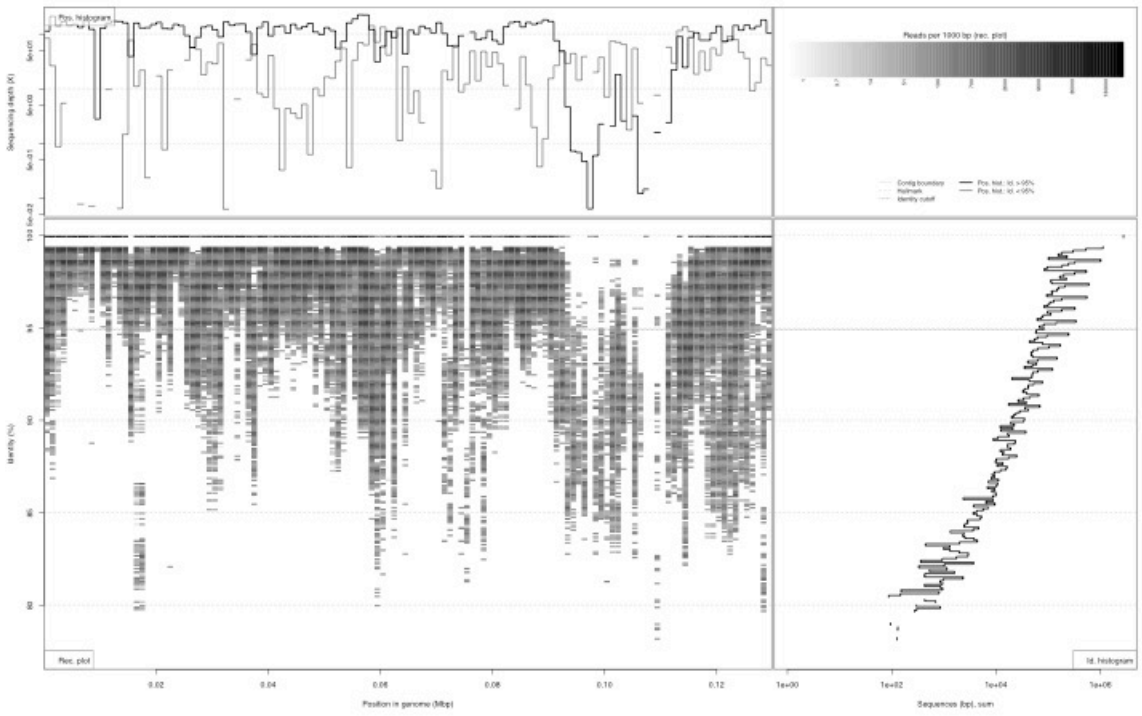
originating from a river sediment (Calcasieu River, USA). The substrates and community ages for each bioreactor are shown. Colors represent the abundant community members, making >1% of the total (see figure key). (B) Schematic representation of the adaptive evolution experimental design. 1% aliquot from the culture that showed a ratio higher than 0.2 of relative growth compared to the control (measured by optical density; red test tubes) for the highest concentration of BAC tested was used as inoculum for the next round with higher BAC concentrations. Yellow tubes denote no growth (due to toxicity by the BAC concentration).



B


Pos. histogram
Reads per 1000 bp (rec. plot)
Seq. plot
Id. histogram
Position in genome (Mbp)
Sequences (bp), sum

C


Pos. histogram
Reads per 1000 bp (rec. plot)
Contig boundary
Pos. hist.: id. >95%
Pos. hist.: id. <95%
Identity cutoff
Sequencing depth (X)
Identity (%)
Rec. plot
Id. histogram
Position in genome (Mbp)
Sequences (bp) sum

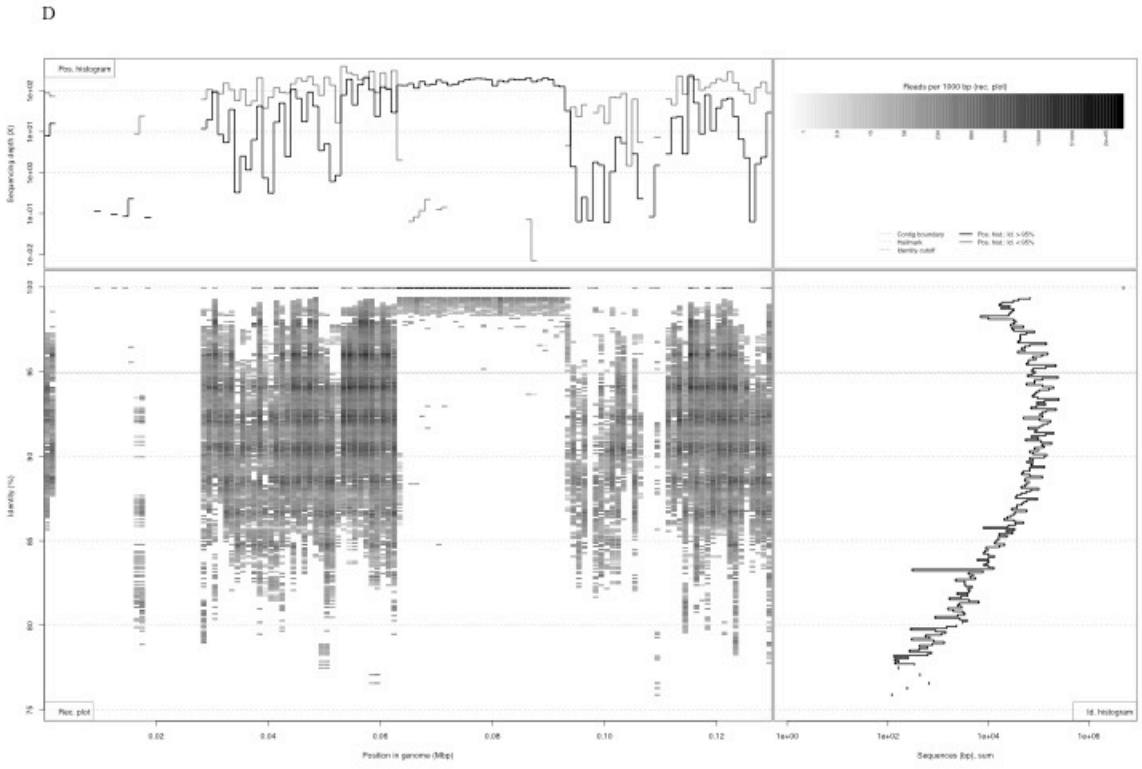


Figure A 2. Fragment recruitment plots of reads of isolate genomes against the *P. aeruginosa* strain DPB GI-2 reference sequence. The fragment recruitment plot consists of four panels: (1) Bottom left panel represents the individual reads recruited against GI-2, placed by location (x-axis) and percent of identity (y-axis) to GI-2 sequence; (2) Top left panel represents sequencing depth across the reference GI-2 sequence, i.e., number of times each nucleotide base of the reference is covered by reads, in logarithmic scale; (3) Bottom right panel represents identity histogram of mapping reads, i.e., how many bases are found at each value of nucleotide identity, in logarithmic scale; and (4) Top right panel represents color scale for the number of stacked reads in bottom left panel. Panels represent reads from: (A) *K. michiganensis* DPB, (B) *Citrobacter freundi* DPB, (C) *Achromobacter sp.* DPB, and (D) *Achromobacter sp.* B.

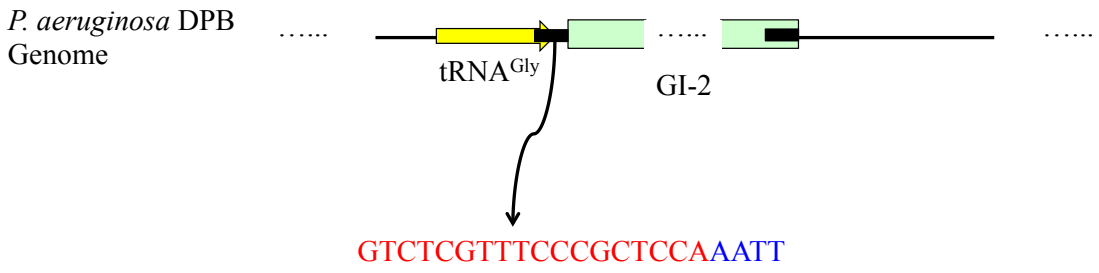


Figure A 3. Schematic representation of the region where the GI-2 element was integrated in the *P. aeruginosa* DPB genome. The yellow arrow indicates the sequence of the $tRNA^{Gly}$ gene. The black rectangles indicate identical short repeated sequences (22 bp) that most likely belong to the *attB* integration site i.e., red-colored nucleotide bases are 100% identical to the tRNA gene while the blue ones do not match the tRNA gene sequence.

Table A 1. Gene content of GI-2.

The table can be found through https://aem.asm.org/content/84/17/e01201-18/figures-only#fig-data-additional-files (Table S1)

Table A 2. MIC values of transformants carrying efflux pump genes

Strain	MIC, mg/L		
	BAC	Tetracycline	Rifampin
PA0509-PBBRMCS4	12.5	0.2	12.5
PA0509/pBBRsugE-A*	25	0.2	12.5
PA0509/pBBRsugE-B	12.5	0.2	12.5
PA0509/pBBRABC	12.5	0.2	25

* PA0509/pBBRsugE-A showed 89.65% reduction in growth at 12.5 mg of BAC/L only once out of three independent replicates. The other two replicates showed ≤ 80% growth reduction at 12.5 mg of BAC/L and ≥ 80% reduction in growth at 25 mg of BAC/L. Hence, 25 mg/L is reported in the Table for BAC.

Table A 3. MIC of antibiotics for *P. aeruginosa* DP & DPB evolved populations. (A) MIC of antibiotics in *P.aeruginosa* DP. (B) MIC of antibiotics in *P.aeruginosa* DPB. (C) MIC of antibiotics supplemented with BAC in *P.aeruginosa* DPB.

A

Populations	MIC, mg/L					
	Tetracycline	Ciprofloxacin	Chloramphenicol	Kanamycin	Rifampin	Ampicillin
P. aeruginosa DP BAC(+)_1	12.5	0.4	25	> 400	25	3200
P. aeruginosa DP BAC(+)_2	12.5	< 0.1	25	50	6.25	400
P. aeruginosa DP BAC(+)_3	12.5	< 0.1	50	100	25	1600
P. aeruginosa DP Control_1	6.25	< 0.1	50	100	25	1600
P. aeruginosa DP BAC(-)_1	6.25	0.025	< 12.5	400	6.25	6400
P. aeruginosa DP BAC(-)_2	3.2	0.1	25	100	< 3.2	100
P. aeruginosa DP BAC(-)_3	12.5	0.8	200	200	12.5	3200
P. aeruginosa DP ancestor	12.5	0.1	50	100	12.5	3200
P. DP						

B

populations	MIC, mg/L						
	Tetracycline	Ciprofloxacin	Chloramphenicol	Polymyxin B	Kanamycin	Rifampin	Ampicillin
P. aeruginosa DPB BAC(+)_1	12.5	0.2	50	0.8	200	25	12800
P. aeruginosa DPB BAC(+)_2	3.125	0.2	25	0.2	50	6.25	400
P. aeruginosa DPB BAC(+)_3	12.5	0.2	50	0.4	100	25	3200
P. aeruginosa DPB BAC(-)_1	6.25	< 0.1	50	0.4	200	25	6400
P. aeruginosa DPB BAC(-)_2	6.25	0.1	12.5	0.4	50	12.5	< 200
P. aeruginosa DPB BAC(-)_3	12.5	0.2	50	0.4	100	25	1600
P. aeruginosa DPB ancestor	6.25	0.2	100	0.2	100	25	3200
P. aeruginosa DPB Control_1	12.5	0.4	100	0.4	100	25	3200
P. aeruginosa DPB Control_2	6.25	0.2	50	0.4	100	12.5	3200

C

populations (LB+BAC cultured)	MIC, mg/L				
	Chloramphenicol +BAC(100mg/L)	Polymyxin B+BAC(100mg/L)	Kanamycin+BAC(100 mg/L)	Rifampin+BAC(100 mg/L)	Ampicillin+BAC(100 mg/L)
P. aeruginosa DPB BAC(+)_1	100	0.8	200	6.25	6400
P. aeruginosa DPB BAC(+)_2	25	0.4	50	6.25	< 200
P. aeruginosa DPB BAC(+)_3	25	0.4	50	< 3.2	3200
P. aeruginosa DPB ancestor	50	0.4	200	< 3.2	6400
P. aeruginosa DPB Control_1	50	0.2	200	< 3.2	6400

Table A 4. (A) Details of SNP mutations *in P. aeruginosa* DP evolved populations. The table can be found through https://aem.asm.org/content/84/17/e01201-18/figures-only#fig-data-additional-files (Table S4A)

(B) **Details of SNP mutations in *P. aeruginosa* DPB evolved populations.** The table can be found through https://aem.asm.org/content/84/17/e01201-18/figures-only#fig-data-additional-files (Table S4B)

(C) **Details of DIP (deletion, insertion and other polymorphisms) mutations in *P. aeruginosa* DP evolved populations.** The table can be found through https://aem.asm.org/content/84/17/e01201-18/figures-only#fig-data-additional-files (Table S4C)

(D) **Details of DIP (deletion, insertion and other polymorphisms) mutations in *P. aeruginosa* DPB evolved populations.** The table can be found through https://aem.asm.org/content/84/17/e01201-18/figures-only#fig-data-additional-files (Table S4D)

Table A 5. Subpopulational mutations in *pmrB* and *pmrA* in *P. aeruginosa* DP BAC(-) populations (Breseq polymorphism mode).

Comparison	Contig	Annotation	Gene position (Contig)	Mutation in contig	Amino acid change	frequency

BAC(-)_2_reads_vs_ancestor_contig	Cont53	*pmrA*	10,036	G→C	L(ancestor and BAC(+)_2) 59 V(BAC(-)_2)	26.80%
		pmrB	8,872	Δ7 bp		7.10%
		pmrB	8,249	C→T	G(ancestor and BAC(+)_2) 421 S(BAC(-)_2)	55.80%
BAC(-)_3_reads_vs_ancestor_contig		*pmrB*	9,449	A→G	F(ancestor and BAC(+)_3) 21 V(BAC(-)_3)	55.20%

Table A 6. (A) Mean transcript abundance and log2 fold change of *arnBCADTEF* under BAC free condition (only LB) in *P. aeruginosa* DPB populations. AL: *P. aeruginosa* DPB ancestor in LB growth media; C2L: *P. aeruginosa* DPB_Control_2 in LB; BAC(+)_1L: *P. aeruginosa* DPB_BAC(+)_1 in LB; BAC(+)_2L: *P. aeruginosa* DPB BAC(+)_2 in LB; BAC(+)_3L: *P. aeruginosa* DPB BAC(+)_3 in LB.

AL_vs_C2L				
	log2FoldChange	**lfcSE**	**baseMean**	**padj**
arnB	-0.19	0.465671379	2881.36	0.95
arnC	-0.39	0.47302977	1451.45	0.86
arnA	-0.42	0.500156202	1981.62	0.85
arnD	-0.13	0.501352141	46.85	0.97
arnT (partial)	-0.13	0.442507489	41.99	0.97
arnT (partial)	-0.33	0.49133792	7.30	0.91
arnE	-0.52	0.404176235	0.62	NA
arnF	-0.95	0.492629556	0.91	NA
AL_vs_BAC(+)_1L				
	log2FoldChange	**lfcSE**	**baseMean**	**padj**
arnB	-0.73	0.431743475	2405.79	0.53
arnC	-1.13	0.472772597	1175.81	0.18
arnA	-0.59	0.504809163	1886.49	0.78
arnD	0.10	0.554740919	52.12	0.99
arnT (partial)	1.01	0.519615554	75.69	0.39
arnT (partial)	1.12	0.524817195	15.96	0.29
arnE	0.30	0.618855137	1.57	NA
arnF	1.04	0.612313641	3.52	0.53
AL_vs_BAC(+)_2L				
	log2FoldChange	**lfcSE**	**baseMean**	**padj**
arnB	0.32	0.533642396	3428.44	0.91
arnC	-0.58	0.569188667	1415.15	0.81
arnA	-0.55	0.570356139	1988.36	0.82
arnD	0.18	0.594907641	52.14	0.97
arnT (partial)	0.58	0.540764554	53.92	0.78
arnT (partial)	0.82	0.584488374	11.75	0.63
arnE	0.01	0.693819434	1.24	1.00
arnF	-1.07	0.71890835	1.07	0.60
AL_vs_BAC(+)_3L				
	log2FoldChange	**lfcSE**	**baseMean**	**padj**
arnB	0.46	0.452365733	3609.00	0.60
arnC	0.27	0.482877747	1820.99	0.81

arnA	0.73	0.506073653	3138.40	0.41
arnD	1.91	0.517494933	128.36	0.00
arnT (partial)	2.24	0.432234197	134.41	0.00
arnT (partial)	2.70	0.456141305	36.24	0.00
arnE	1.20	0.809027602	2.77	0.39
arnF	1.38	0.728107695	3.82	0.23

(B) Mean transcript abundance and log2 fold change of *arnBCADTEF* under BAC exposure (LB+BAC) in *P. aeruginosa* DPB populations. AB: *P. aeruginosa* DPB ancestor in LB+BAC media; C2B: *P. aeruginosa* DPB Control_2 in LB+BAC; BAC(+)_1B: *P. aeruginosa* DPB BAC(+)_1 in LB+BAC; BAC(+)_2B: *P. aeruginosa* DPB BAC(+)_2 in LB+BAC; BAC(+)_3B: *P. aeruginosa* DPB BAC(+)_3 in LB+BAC.

AB_vs_C2B				
	log2FoldChange	**lfcSE**	**baseMean**	**padj**
arnB	0.09	0.215104554	801.92	1.00
arnC	0.12	0.23495053	550.45	1.00
arnA	0.01	0.207432747	1599.01	1.00
arnD	-0.20	0.327154778	43.52	1.00
arnT (partial)	-0.01	0.302074495	67.17	1.00
arnT (partial)	-0.59	0.415935123	12.70	0.84
arnE	0.31	0.404163246	0.97	NA
arnF	0.12	0.453587471	1.20	NA
AB_vs_BAC(+)_1B				
	log2FoldChange	**lfcSE**	**baseMean**	**padj**
arnB	1.28	0.23754236	1160.45	0.00
arnC	1.36	0.269060725	824.93	0.00
arnA	1.45	0.224670717	2606.73	0.00
arnD	1.14	0.31840902	65.64	0.00
arnT (partial)	1.79	0.304537468	134.99	0.00
arnT (partial)	0.79	0.435402146	19.35	0.24
arnE	0.49	0.730632955	0.87	NA
arnF	1.48	0.727945525	2.56	NA
AB_vs_BAC(+)_2B				
	log2FoldChange	**lfcSE**	**baseMean**	**padj**
arnB	1.00	0.294281292	1088.14	0.01
arnC	1.24	0.236915578	825.38	0.00
arnA	1.14	0.297513844	2395.02	0.00
arnD	1.07	0.395549988	69.30	0.05
arnT (partial)	1.61	0.315054698	130.25	0.00
arnT (partial)	0.56	0.396600845	18.63	0.45

arnE	-0.13	0.586931586	0.37	NA
arnF	0.82	0.73034847	1.74	0.59
AB_vs_BAC(+)_3B				
	log2FoldChange	**lfcSE**	**baseMean**	**padj**
arnB	2.90	0.21317264	2516.05	0.00
arnC	3.15	0.195491069	1989.37	0.00
arnA	2.83	0.216985234	4967.34	0.00
arnD	2.72	0.250158093	136.69	0.00
arnT (partial)	2.88	0.241083376	218.91	0.00
arnT (partial)	1.94	0.368403034	30.47	0.00
arnE	1.84	0.828393669	1.90	NA
arnF	2.22	0.767476884	3.86	0.02

Table A 7. Statistics of RNA-seq libraries. The table can be found through https://aem.asm.org/content/84/17/e01201-18/figures-only#fig-data-additional-files (Table S7)

APPENDIX B SUPPLEMENTARY MATERIAL FOR CHAPTER 3

B.1 Supplementary figures and tables

Table B 1. (A) Details of SNP mutations found in *P. aeruginosa* DP evolved populations. The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S1A)

(B) Details of SNP mutations in *P. aeruginosa* DPB evolved populations. The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S1B)

(C) Details of DIP (deletion, insertion and other polymorphisms) mutations detected in *P. aeruginosa* DP evolved populations. The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S1C)

(D) Details of DIP (deletion, insertion and other polymorphisms) mutations detected in *P. aeruginosa* DPB evolved populations. The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S1D)

(E) Mutations in *pmrB* and *pmrA* genes found in *P. aeruginosa* DP BAC(-) populations (Breseq polymorphism mode). The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S1E)

Table B 2.

The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S2)

Table B 3.

The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S3)

Table B 4.

The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S4)

Table B 5.

The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S5)

Table B 6.

The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S6)

Table B 7.

The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S7)

Table B 8.

The table can be found through https://aem.asm.org/content/84/12/e00197-18/figures-only#fig-data-additional-files (Table S8)

APPENDIX C SUPPLEMENTARY MATERIAL FOR CHAPTER 4

C.1 Supplementary figures and tables

Table C 1. Taxonomic classification of recovered MAGs. The Microbial Genome Atlas (MiGA) webserver was used to determine the most likely taxonomic classification of MAGs against the classified species in NCBI's prokaryotic genome database (222).

MAG	Species_ID	Closest relative and accession ID	AAI (%)
Map_122_maxbin_012	ANIsp_1	*Escherichia_coli*_str__K_12_substr__MDS42_NC_020518	99.22
Map_49_maxbin_001	ANIsp_2	*Prevotella_jejuni*_CP023863	62.19
Map_3_maxbin_001	ANIsp_3	*Prevotella_jejuni*_CP023863	62.10
Map_16_maxbin_009	ANIsp_4	*Bifidobacterium_longum*_subsp__infantis_NZ_CP010411	97.72
Map_67_maxbin_001	ANIsp_5	*Roseburia_hominis*_A2_183_NC_015977	41.96
Map_106_maxbin_016	ANIsp_6	*Veillonella_atypica*_CP020566	61.30
Map_66_maxbin_003	ANIsp_7	*Bacteroides_vulgatus*_ATCC_8482_NC_009614	97.64
Map_55_maxbin_002	ANIsp_8	*Bifidobacterium_kashiwanohense*_PV20_2_NZ_CP007456	95.84
Map_41_maxbin_004	ANIsp_9	*Bacteroides_fragilis*_NZ_LN877293	98.24
Map_114_maxbin_008	ANIsp_10	*Prevotella_jejuni*_CP023863	61.26
Map_20_maxbin_016	ANIsp_11	*Bifidobacterium_bifidum*_S17_NC_014616	98.04
Map_109_maxbin_010	ANIsp_12	*Dialister_pneumosintes*_NZ_CP017037	56.18
Map_9_maxbin_004	ANIsp_13	*Bifidobacterium_breve*_CP021387	98.07
Map_75_maxbin_009	ANIsp_14	*Prevotella_enoeca*_NZ_CP013195	61.91
Map_85_maxbin_015	ANIsp_15	*_Eubacterium__rectale*_M104_1_FP929043	96.19
Map_113_maxbin_021	ANIsp_16	*Libanicoccus_massiliensis*_NZ_LT671675	65.73
Map_118_maxbin_025	ANIsp_17	*Prevotella_enoeca*_NZ_CP013195	61.93
Map_39_maxbin_007	ANIsp_18	*Ruminococcaceae_bacterium*_CPB6_NZ_CP020705	48.14
Map_112_maxbin_002	ANIsp_19	*Prevotella_dentalis*_DSM_3688_NC_019960	61.13
Map_91_maxbin_009	ANIsp_20	*Prevotella_enoeca*_NZ_CP013195	61.53
Map_30_maxbin_015	ANIsp_21	*Megamonas_hypermegale*_NZ_LT906446	80.22
Map_100_maxbin_009	ANIsp_22	*Pasteurella_multocida*_CP023304	47.25

Map_14_maxbin_019	ANIsp_23	*Bacteroides_vulgatus*_NZ_CP013020	56.43
Map_106_maxbin_008	ANIsp_24	*Collinsella_aerofaciens*_NZ_CP024160	92.66
Map_8_maxbin_012	ANIsp_25	*Parabacteroides_distasonis*_ATCC_8503_NC_009615	74.11
Map_114_maxbin_012	ANIsp_26	*Bacteroides_vulgatus*_NZ_CP013020	56.79
Map_51_maxbin_007	ANIsp_27	*Muribaculum_intestinale*_NZ_CP021421	56.49
Map_57_maxbin_005	ANIsp_28	*Bacteroides_ovatus*_V975_NZ_LT622246	94.76
Map_111_maxbin_039	ANIsp_29	*Akkermansia_muciniphila*_ATCC_BAA_835_NC_010655	97.80
Map_66_maxbin_001	ANIsp_30	*Parabacteroides*_sp__CT06_NZ_CP022754	95.47
Map_23_maxbin_012	ANIsp_31	*Streptococcus*_sp__HSISS3_CM002130	96.72
Map_42_maxbin_008	ANIsp_32	*Bacteroides_fragilis*_CP018937	96.95
Map_12_maxbin_013	ANIsp_33	*Alistipes_finegoldii*_DSM_17242_NC_018011	66.79
Map_96_maxbin_002	ANIsp_34	*Prevotella_jejuni*_CP023863	60.63
Map_91_maxbin_024	ANIsp_35	*Prevotella_denticola*_F0289_NC_015311	63.19
Map_113_maxbin_023	ANIsp_36	*Lactobacillus_mucosae*_LM1_NZ_CP011013	96.61
Map_7_maxbin_003	ANIsp_37	*Bacteroides_helcogenes*_P_36_108_NC_014933	78.84
Map_102_maxbin_006	ANIsp_38	*Turicibacter*_sp__H121_NZ_CP013476	40.75
Map_77_maxbin_005	ANIsp_39	*Bacteroides_thetaiotaomicron*_VPI_5482_NC_004663	98.13
Map_107_maxbin_013	ANIsp_40	*_Eubacterium__rectale*_ATCC_33656_NC_012781	64.78
Map_24_maxbin_011	ANIsp_41	*Prevotella_enoeca*_NZ_CP013195	61.79
Map_10_maxbin_003	ANIsp_42	*Veillonella_parvula*_NZ_CP019721	90.93
Map_121_maxbin_015	ANIsp_43	*Bacteroides_caccae*_NZ_CP022412	98.43
Map_118_maxbin_028	ANIsp_44	*Lactobacillus_ruminis*_ATCC_27782_NC_015975	96.57
Map_67_maxbin_007	ANIsp_45	*Ruminococcus_torques*_L2_14_FP929055	66.52
Map_7_maxbin_013	ANIsp_46	*Faecalibacterium_prausnitzii*_SL3_3_FP929046	52.42
Map_75_maxbin_018	ANIsp_47	*Bacteroides_vulgatus*_NZ_CP013020	56.74
Map_103_maxbin_029	ANIsp_48	*Bacteroidales_bacterium*_CF_NZ_CP006772	50.56
Map_58_maxbin_005	ANIsp_49	*Anaerostipes_hadrus*_NZ_CP012098	97.10
Map_52_maxbin_023	ANIsp_50	*Bacteroides_helcogenes*_P_36_108_NC_014933	77.96
Map_4_maxbin_006	ANIsp_51	*Streptococcus_infantarius*_NZ_CP013689	98.93
Map_92_maxbin_020	ANIsp_52	*Acidaminococcus_fermentans*_DSM_20731_NC_013740	53.08
Map_119_maxbin_013	ANIsp_53	*Streptococcus_gallolyticus*_subsp__gallolyticus_ATCC_43143_NC_017576	94.28
Map_14_maxbin_010	ANIsp_54	*Prevotella_jejuni*_CP023863	60.54

Map_68_maxbin_005	ANIsp_55	*Erysipelotrichaceae_bacterium*_I46_NZ_CP015404	45.41
Map_105_maxbin_003	ANIsp_56	*Faecalibacterium_prausnitzii*_NZ_CP022479	85.18
Map_102_maxbin_015	ANIsp_57	*Brachyspira_pilosicoli*_P43_6_78_NC_019908	97.94
Map_103_maxbin_024	ANIsp_58	*Lactococcus_garvieae*_ATCC_49156_NC_015930	95.63
Map_58_maxbin_010	ANIsp_59	*Faecalibacterium_prausnitzii*_L2_6_FP929045	56.67
Map_105_maxbin_029	ANIsp_60	*Streptococcus_pasteurianus*_ATCC_43144_NC_015600	98.14
Map_52_maxbin_018	ANIsp_61	*Barnesiella_viscericola*_DSM_18177_NZ_CP007034	80.30
Map_109_maxbin_019	ANIsp_62	*Ruminococcaceae_bacterium*_CPB6_NZ_CP020705	47.64
Map_74_maxbin_008	ANIsp_63	*Erysipelotrichaceae_bacterium*_I46_NZ_CP015404	45.24
Map_119_maxbin_014	ANIsp_64	*Megasphaera_elsdenii*_DSM_20460_NC_015873	93.40
Map_3_maxbin_005	ANIsp_65	*Prevotella_enoeca*_NZ_CP013195	62.38
Map_77_maxbin_020	ANIsp_66	*Veillonella_parvula*_NZ_LT906445	61.35
Map_85_maxbin_011	ANIsp_67	*Faecalibacterium_prausnitzii*_NZ_CP022479	81.01
Map_72_maxbin_007	ANIsp_68	*Megasphaera_elsdenii*_DSM_20460_NC_015873	81.16
Map_85_maxbin_026	ANIsp_69	*Acutalibacter_muris*_NZ_CP021422	50.09
Map_78_maxbin_010	ANIsp_70	*Bacteroides_xylanisolvens*_XB1A_FP929033	96.07
Map_98_maxbin_013	ANIsp_71	*Acidaminococcus_fermentans*_DSM_20731_NC_013740	53.06
Map_102_maxbin_009	ANIsp_72	*Faecalibacterium_prausnitzii*_NZ_CP022479	79.13
Map_106_maxbin_004	ANIsp_73	*Mycoplasma_suis*_KI3806_NC_015153	40.53
Map_105_maxbin_005	ANIsp_74	*Bifidobacterium_pseudocatenulatum*_DSM_20438___JCM_1200___LMG_10505_NZ_AP012330	97.94
Map_77_maxbin_002	ANIsp_75	*Ruminococcus*_sp__SR1_5_FP929053	68.60
Map_15_maxbin_014	ANIsp_76	*Oscillibacter_valericigenes*_Sjm18_20_NC_016048	56.41
Map_30_maxbin_006	ANIsp_77	*Haemophilus_parainfluenzae*_T3T1_NC_015964	96.76
Map_35_maxbin_020	ANIsp_78	*Campylobacter_fetus*_subsp__venerealis_NCTC_10354_NZ_CM001228	55.62
Map_82_maxbin_008	ANIsp_79	*Streptococcus_thermophilus*_ND03_NC_017563	98.68
Map_85_maxbin_025	ANIsp_80	*Streptococcus_salivarius*_NZ_CP015282	97.01
Map_100_maxbin_003	ANIsp_81	*Prevotella_dentalis*_DSM_3688_NC_019	61.71

		960	
Map_102_maxbin_014	ANIsp_82	*Collinsella_aerofaciens*_NZ_CP024160	41.64
Map_102_maxbin_017	ANIsp_83	*Roseburia_intestinalis*_M50_1_FP929049	53.55
Map_102_maxbin_022	ANIsp_84	*Blautia_coccoides*_NZ_CP022713	68.02
Map_102_maxbin_024	ANIsp_85	*_Clostridium__innocuum*_NZ_CP022722	46.45
Map_103_maxbin_023	ANIsp_86	*Candidatus*_Hodgkinia_cicadicola_CP024746	40.44
Map_103_maxbin_031	ANIsp_87	*Elusimicrobium_minutum*_Pei191_NC_010644	52.09
Map_104_maxbin_012	ANIsp_88	*Dialister_pneumosintes*_NZ_CP017037	56.06
Map_105_maxbin_006	ANIsp_89	*Blautia_obeum*_A2_162_FP929054	80.45
Map_105_maxbin_031	ANIsp_90	*Bacteroides_salanitronis*_DSM_18170_NC_015164	74.82
Map_105_maxbin_033	ANIsp_91	*Acutalibacter_muris*_NZ_CP021422	50.74
Map_107_maxbin_018	ANIsp_92	*Dialister_pneumosintes*_NZ_CP017037	56.08
Map_110_maxbin_009	ANIsp_93	*Paeniclostridium_sordellii*_CP014150	40.50
Map_111_maxbin_016	ANIsp_94	*Roseburia_hominis*_A2_183_NC_015977	68.02
Map_111_maxbin_024	ANIsp_95	*Prevotella_dentalis*_DSM_3688_NC_019960	61.94
Map_115_maxbin_017	ANIsp_96	*Roseburia_intestinalis*_M50_1_FP929049	53.33
Map_115_maxbin_030	ANIsp_97	*_Clostridium__cellulosi*_NZ_LM995447	47.36
Map_117_maxbin_021	ANIsp_98	*_Eubacterium__hallii*_NZ_LT907978	41.33
Map_117_maxbin_025	ANIsp_99	*Parageobacillus_genomosp__*1_NZ_CM002692	42.80
Map_117_maxbin_026	ANIsp_100	*Ruminiclostridium*_sp__KB18_NZ_CP015400	48.62
Map_118_maxbin_024	ANIsp_101	*Bacteroidales_bacterium*_CF_NZ_CP006772	50.67
Map_118_maxbin_026	ANIsp_102	*Candidatus*_Hodgkinia_cicadicola_CP024746	40.13
Map_118_maxbin_027	ANIsp_103	*Bacteroides_vulgatus*_NZ_CP013020	53.83
Map_118_maxbin_029	ANIsp_104	*Ruminococcaceae_bacterium*_CPB6_NZ_CP020705	48.01
Map_120_maxbin_004	ANIsp_105	*Prevotella_dentalis*_DSM_3688_NC_019960	60.56
Map_122_maxbin_020	ANIsp_106	*Burkholderiales_bacterium*_YL45_NZ_CP015403	52.87
Map_14_maxbin_012	ANIsp_107	*Faecalibaculum_rodentium*_NZ_CP011391	55.21
Map_24_maxbin_010	ANIsp_108	*Bacteroides_vulgatus*_NZ_CP013020	56.34
Map_25_maxbin_008	ANIsp_109	*Akkermansia_muciniphila*_ATCC_BAA_835_NC_010655	88.45
Map_27_maxbin_007	ANIsp_110	*Morganella_morganii*_NZ_CP026046	97.90
Map_27_maxbin_009	ANIsp_111	*Lactobacillus_gasseri*_ATCC_33323___JCM_1131_NC_008530	97.99
Map_27_maxbin_010	ANIsp_112	*Lactobacillus_salivarius*_str__Ren_NZ_C	97.66

		P011403	
Map_35_maxbin_011	ANIsp_113	*Helicobacter_mustelae*_12198_NC_013949	52.94
Map_39_maxbin_019	ANIsp_114	*Collinsella_aerofaciens*_NZ_CP024160	61.08
Map_46_maxbin_012	ANIsp_115	*Candidatus*_Melainabacteria_bacterium_MEL_A1_CP017245	59.82
Map_47_maxbin_006	ANIsp_116	*Burkholderiales_bacterium*_YL45_NZ_CP015403	62.12
Map_4_maxbin_007	ANIsp_117	*Bifidobacterium_longum*_subsp__infantis_NZ_CP010411	92.90
Map_51_maxbin_023	ANIsp_118	*Christensenella_massiliensis*_NZ_LT700187	45.16
Map_52_maxbin_022	ANIsp_119	*Megasphaera_elsdenii*_DSM_20460_NC_015873	97.14
Map_54_maxbin_004	ANIsp_120	*Blautia_obeum*_A2_162_FP929054	67.20
Map_55_maxbin_011	ANIsp_121	*Enterobacter_cloacae*_NZ_CP019839	97.83
Map_56_maxbin_006	ANIsp_122	*Bifidobacterium_adolescentis*_NZ_CP007443	97.35
Map_57_maxbin_008	ANIsp_123	*Collinsella_aerofaciens*_NZ_CP024160	88.10
Map_57_maxbin_013	ANIsp_124	*Akkermansia_muciniphila*_NZ_CP021420	88.65
Map_58_maxbin_008	ANIsp_125	*Flavonifractor_plautii*_NZ_CP015406	49.35
Map_58_maxbin_015	ANIsp_126	*Clostridium__bolteae*_NZ_CP022464	64.70
Map_58_maxbin_020	ANIsp_127	*Clostridium_chauvoei*_NZ_CP018624	63.78
Map_67_maxbin_004	ANIsp_128	*Flavonifractor_plautii*_NZ_CP015406	48.93
Map_67_maxbin_006	ANIsp_129	butyrate_producing_bacterium_SS3_4_FP929062	86.65
Map_67_maxbin_008	ANIsp_130	*Veillonella_parvula*_NZ_LT906445	95.27
Map_67_maxbin_012	ANIsp_131	*Lachnoclostridium_phocaeense*_NZ_LT635479	63.14
Map_67_maxbin_014	ANIsp_132	*Alistipes_finegoldii*_DSM_17242_NC_018011	82.10
Map_69_maxbin_011	ANIsp_133	*Bacteroides_vulgatus*_NZ_CP013020	56.34
Map_6_maxbin_010	ANIsp_134	*Clostridium__cellulosi*_NZ_LM995447	46.88
Map_6_maxbin_015	ANIsp_135	*Acidaminococcus_fermentans*_DSM_20731_NC_013740	53.52
Map_6_maxbin_017	ANIsp_136	*Enterococcus_hirae*_ATCC_9790_NC_018081	97.73
Map_74_maxbin_009	ANIsp_137	*Faecalibaculum_rodentium*_NZ_CP011391	55.35
Map_74_maxbin_018	ANIsp_138	*Campylobacter_coli*_NZ_CP015528	58.67
Map_75_maxbin_013	ANIsp_139	*Faecalibaculum_rodentium*_NZ_CP011391	54.76
Map_77_maxbin_019	ANIsp_140	*Megasphaera_elsdenii*_14_14_NZ_CP009240	61.96
Map_81_maxbin_026	ANIsp_141	*Dialister_pneumosintes*_NZ_CP017037	56.03
Map_82_maxbin_005	ANIsp_142	*Collinsella_aerofaciens*_NZ_CP024160	90.90

Map_83_maxbin_007	ANIsp_143	*Streptococcus_parasanguinis*_ATCC_15912_NC_015678	46.08
Map_83_maxbin_010	ANIsp_144	*Atopobium_parvulum*_DSM_20469_NC_013203	92.12
Map_88_maxbin_002	ANIsp_145	*Prevotella_dentalis*_DSM_3688_NC_019960	59.56
Map_8_maxbin_008	ANIsp_146	*Odoribacter_splanchnicus*_NZ_LT906459	98.04
Map_92_maxbin_013	ANIsp_147	*Prevotella_jejuni*_CP023863	59.72
Map_92_maxbin_016	ANIsp_148	*Collinsella_aerofaciens*_NZ_CP024160	90.17
Map_96_maxbin_010	ANIsp_149	*Oscillibacter_valericigenes*_Sjm18_20_NC_016048	55.71
Map_98_maxbin_018	ANIsp_150	*Oscillibacter_valericigenes*_Sjm18_20_NC_016048	56.95
Map_98_maxbin_031	ANIsp_151	*Roseburia_hominis*_A2_183_NC_015977	53.64
Map_98_maxbin_033	ANIsp_152	*Alistipes_finegoldii*_DSM_17242_NC_018011	97.61
Map_9_maxbin_003	ANIsp_153	*Klebsiella_pneumoniae*_CP011421	99.10

Table C 2. Permutational multivariate analysis of variance (PERMANOVA) tests of metadata and Mash distances among samples (permutation = 999, Significance codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1).**

Metadata	**R^2**	**Pr(>F)**	**sig**
age_sampCat	0.31518	0.001	***
Studyarm_binary	0.01239	0.184	
female	0.01246	0.181	
Compound_sanitary_index	0.0145	0.127	
Ppl_per_Hhroom_categorical	0.01755	0.077	.
HL_populationdensity_quintiles	0.01357	0.156	
chronpain	0.01092	0.257	
diarrhea	0.02274	0.042	*
age_sampCat:Studyarm_binary	0.02186	0.243	
age_sampCat:female	0.01969	0.292	
Studyarm_binary:female	0.01628	0.103	
age_sampCat:Compound_sanitary_index	0.02222	0.236	
Studyarm_binary:Compound_sanitary_index	0.00617	0.54	
female:Compound_sanitary_index	-0.003	0.999	
age_sampCat:Ppl_per_Hhroom_categorical	0.01443	0.488	
Studyarm_binary:Ppl_per_Hhroom_categorical	0.00666	0.494	
female:Ppl_per_Hhroom_categorical	0.00445	0.714	
Compound_sanitary_index:Ppl_per_Hhroom_categorical	0.00223	0.914	
age_sampCat:HL_populationdensity_quintiles	0.01102	0.693	
Studyarm_binary:HL_populationdensity_quintiles	0.01162	0.233	
female:HL_populationdensity_quintiles	0.00344	0.804	

Compound_sanitary_index:HL_populationdensity_quintiles	0.00329	0.837	
Ppl_per_Hhroom_categorical:HL_populationdensity_quintiles	0.01073	0.254	
age_sampCat:chronpain	0.01488	0.508	
Studyarm_binary:chronpain	0.0054	0.629	
female:chronpain	0.00983	0.259	
age_sampCat:diarrhea	0.0114	0.652	
Studyarm_binary:diarrhea	0.00517	0.654	
female:diarrhea	0.00313	0.841	
Compound_sanitary_index:diarrhea	0.02775	0.017	*
Ppl_per_Hhroom_categorical:diarrhea	0.01193	0.191	
HL_populationdensity_quintiles:diarrhea	0.00773	0.435	
age_sampCat:Studyarm_binary:female	0.01019	0.75	
age_sampCat:Studyarm_binary:Compound_sanitary_index	0.03359	0.056	.
age_sampCat:female:Compound_sanitary_index	0.01511	0.472	
Studyarm_binary:female:Compound_sanitary_index	0.0059	0.604	
age_sampCat:Studyarm_binary:Ppl_per_Hhroom_categorical	0.01185	0.21	
age_sampCat:female:Ppl_per_Hhroom_categorical	0.00493	0.658	
Studyarm_binary:female:Ppl_per_Hhroom_categorical	0.00972	0.306	
age_sampCat:Compound_sanitary_index:Ppl_per_Hhroom_categorical	0.02757	0.119	
Studyarm_binary:Compound_sanitary_index:Ppl_per_Hhroom_categorical	0.00859	0.369	
female:Compound_sanitary_index:Ppl_per_Hhroom_categorical	0.00485	0.657	
age_sampCat:Studyarm_binary:HL_populationdensity_quintiles	0.02184	0.056	.
age_sampCat:female:HL_populationdensity_quintiles	0.00272	0.886	
Studyarm_binary:female:HL_populationdensity_quintiles	0.00784	0.416	
age_sampCat:Compound_sanitary_index:HL_populationdensity_quintiles	0.00945	0.302	
Studyarm_binary:Compound_sanitary_index:HL_populationdensity_quintiles	0.01211	0.193	
female:Compound_sanitary_index:HL_populationdensity_quintiles	0.00319	0.852	
age_sampCat:Ppl_per_Hhroom_categorical:HL_populationdensity_quintiles	0.01316	0.167	
Studyarm_binary:Ppl_per_Hhroom_categorical:HL_populationdensity_quintiles	0.00774	0.424	
female:Ppl_per_Hhroom_categorical:HL_populationdensity_quintiles	0.00346	0.8	
age_sampCat:HL_populationdensity_quintiles:diarrhea	0.01254	0.175	
age_sampCat:Studyarm_binary:Compound_sanitary_index:HL_populationdensity_quintiles	0.01166	0.21	
age_sampCat:female:Compound_sanitary_index:HL_populationdensity_quintiles	0.00589	0.564	
Residuals	0.08856		
Total	1		

Table C 3 Permutational multivariate analysis of variance (PERMANOVA) tests of metadata and Mash distances among samples in the 1-11 months old group (permutation = 999, Significance codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1).**

Metadata	R^2	Pr(>F)	sig
Studyarm_binary	0.04529	0.083	.
female	0.04308	0.116	
Compound_sanitary_index	0.02693	0.445	
Ppl_per_Hhroom_categorical	0.04138	0.15	
HL_populationdensity_quintiles	0.02181	0.621	
chronpain	0.01698	0.82	
diarrhea	0.02859	0.398	
Studyarm_binary:female	0.02879	0.373	
Studyarm_binary:Compound_sanitary_index	0.04205	0.139	
female:Compound_sanitary_index	0.01406	0.911	
Studyarm_binary:Ppl_per_Hhroom_categorical	0.05145	0.054	.
female:Ppl_per_Hhroom_categorical	0.03011	0.373	
Compound_sanitary_index:Ppl_per_Hhroom_categorical	0.02786	0.413	
Studyarm_binary:HL_populationdensity_quintiles	0.0306	0.32	
female:HL_populationdensity_quintiles	0.02339	0.556	
Compound_sanitary_index:HL_populationdensity_quintiles	0.03834	0.175	
Ppl_per_Hhroom_categorical:HL_populationdensity_quintiles	0.02841	0.393	
female:chronpain	0.02639	0.428	
Studyarm_binary:diarrhea	0.01657	0.87	
female:diarrhea	0.01231	0.953	
Ppl_per_Hhroom_categorical:diarrhea	0.03399	0.258	
HL_populationdensity_quintiles:diarrhea	0.05585	0.051	.
Studyarm_binary:female:Compound_sanitary_index	0.01898	0.741	
Studyarm_binary:female:HL_populationdensity_quintiles	0.03368	0.258	
Studyarm_binary:Compound_sanitary_index:HL_populationdensity_quintiles	0.02591	0.465	
female:Compound_sanitary_index:HL_populationdensity_quintiles	0.01803	0.792	
Residuals	0.21918		
Total	1		

Table C 4 Permutational multivariate analysis of variance (PERMANOVA) tests of metadata and Mash distances among samples in the 24-48 months old group (permutation = 999, Significance codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1).**

Metadata	R^2	Pr(>F)	sig
Studyarm_binary	0.0406	0.071	.
female	0.0359	0.122	

Compound_sanitary_index	0.04438	0.044	*
Ppl_per_Hhroom_categorical	0.06091	0.002	**
HL_populationdensity_quintiles	0.05438	0.008	**
chronpain	0.0216	0.715	
diarrhea	0.04822	0.023	*
Studyarm_binary:female	0.03545	0.139	
Studyarm_binary:Compound_sanitary_index	0.04315	0.053	.
female:Compound_sanitary_index	0.03149	0.222	
Studyarm_binary:Ppl_per_Hhroom_categorical	0.018	0.858	
female:Ppl_per_Hhroom_categorical	0.02854	0.36	
Compound_sanitary_index:Ppl_per_Hhroom_categorical	0.03526	0.135	
Studyarm_binary:HL_populationdensity_quintiles	0.03454	0.169	
female:HL_populationdensity_quintiles	0.0293	0.33	
Compound_sanitary_index:HL_populationdensity_quintiles	0.02757	0.389	
Ppl_per_Hhroom_categorical:HL_populationdensity_quintiles	0.03391	0.153	
Studyarm_binary:diarrhea	0.04279	0.047	*
female:diarrhea	0.02765	0.355	
HL_populationdensity_quintiles:diarrhea	0.03407	0.191	
Studyarm_binary:female:Compound_sanitary_index	0.04553	0.034	*
Studyarm_binary:female:Ppl_per_Hhroom_categorical	0.03055	0.282	
Studyarm_binary:Compound_sanitary_index:Ppl_per_Hhroom_categorical	0.02932	0.283	
Studyarm_binary:female:HL_populationdensity_quintiles	0.03212	0.227	
Studyarm_binary:Compound_sanitary_index:HL_populationdensity_quintiles	0.03326	0.202	
female:Compound_sanitary_index:HL_populationdensity_quintiles	0.02292	0.636	
Residuals	0.0786		
Total	1		

Table C 5. *E. coli* diagnostic genes used in this study to assess virulence potential of recovered *E. coli* MAGs.

Pathotype	Diagnostic gene accession ID	Annotation
ETEC	>VFG002037(gb|AAA24685)_(eltA)_heat-labile_enterotoxin_A_prepeptide_(from_human)_[Heat-labile_toxin_(LT)_(VF0210)]_[Escherichia_coli]	Heat-labile enterotoxin (subunit A)
ETEC	>VFG002038(gb|AAA98064)_(eltB)_enterotoxin_subunit_B_(from_human)_[Heat-labile_toxin_(LT)_(VF0210)]_[Escherichia_coli]	Heat-labile enterotoxin (subunit B)
ETEC	>VFG002039(gb|YP_003294006)_(estIa)_heat_stable_enterotoxin_I_[Heat-stable_toxin_(ST)_(VF0211)]_[Escherichia_coli_O78:H11:K80_str._H10407] >VFG000863(gb|BAA94855)_(astA)_heat-stable_enterotoxin_1_[EAST1_(VF0216)]_[Escherichia_coli_O44:H18_042]	Heat-stable enterotoxin

EPEC(typical)	>VFG000749(gb|BAA84838)_(bfpA)_Bundlin_[BFP_(VF0174)]_[Escherichia_coli_B171]	Bundle-forming pilus structural gene A
EPEC(atypical)	>VFG000739(gb|AAC38392)_(eae)_intimin_[Intimin_(VF0177)]_[Escherichia_coli_O127:H6_str._E2348/69]	Gamma Intimin
EHEC	>VFG000835(gb|NP_288673)_(stx1A)_shiga-like_toxin_1_subunit_A_encoded_within_prophage_CP-933V_[Stx_(VF0206)]_[Escherichia_coli_O157:H7_str._EDL933]	Shiga-toxin (subunit A)
EHEC	>VFG000836(gb|NP_288672)_(stx1B)_shiga-like_toxin_1_subunit_B_encoded_within_prophage_CP-933V_[Stx_(VF0206)]_[Escherichia_coli_O157:H7_str._EDL933]	Shiga-toxin (subunit B)
EAEC	>VFG000851(gb|AAB82330)_(aafA)_major_fimbrial_subunit_of_aggregative_adherence_fimbria_II_AafA_[AAFs_(VF0214)]_[Escherichia_coli_str._042] >VFG000853(gb|AAD27810)_(aafC)_aggregative_adherence_fimbria_II_usher_protein_AafC_[AAFs_(VF0214)]_[Escherichia_coli_str._042] >VFG000855(gb|AAM88298)_(agg3A)_adhesin_protein_[AAFs_(VF0214)]_[Escherichia_coli_55989] >VFG000856(gb|AAM88297)_(agg3B)_invasin_protein_[AAFs_(VF0214)]_[Escherichia_coli_55989]	aggregative adherence fimbria II, aggregatvie adherence fimbria II, adhesion protein, and invasin protein
DAEC	>VFG042424(gb|CAW30798)_(afaB-I)_chaperone_AfaB_[Afimbrial_adhesin,_AFA-I,_mannose-resistant_adhesin_(AI010)]_[Escherichia_coli_O25b:H4_str._FV9863]VFG042429(gb|CAA54112)_(afaF-III)_AfaF-III_[Afimbrial_adhesin,_AFA-III_(AI012)]_[Escherichia_coli_str._A30] >VFG042425(gb|CAW30799)_(afaC-I)_usher_AfaC_[Afimbrial_adhesin,_AFA-I,_mannose-resistant_adhesin_(AI010)]_[Escherichia_coli_O25b:H4_str._FV9863] >VFG042423(gb|CAW30797)_(afaA)_fimbrial_major_subunit_AfaA_[Afimbrial_adhesin,_AFA-I,_mannose-resistant_adhesin_(AI010)]_[Escherichia_coli_O25b:H4_str._FV9863] >VFG001426(gb|CAW30801)_(afaE-I)_AFA-I_adhesin_[Afa/Dr_family_(VF0212)]_[Escherichia_coli_O25b:H4_str._FV9863] >VFG001425(gb|CAA54121)_(afaE-III)_AFA-III_adhesin_[Afa/Dr_family_(VF0212)]_[Escherichia_coli_str._A30] >VFG042429(gb|CAA54112)_(afaF-III)_AfaF-III_[Afimbrial_adhesin,_AFA-III_(AI012)]_[Escherichia_coli_str._A30]	afimbrial adhesin operon

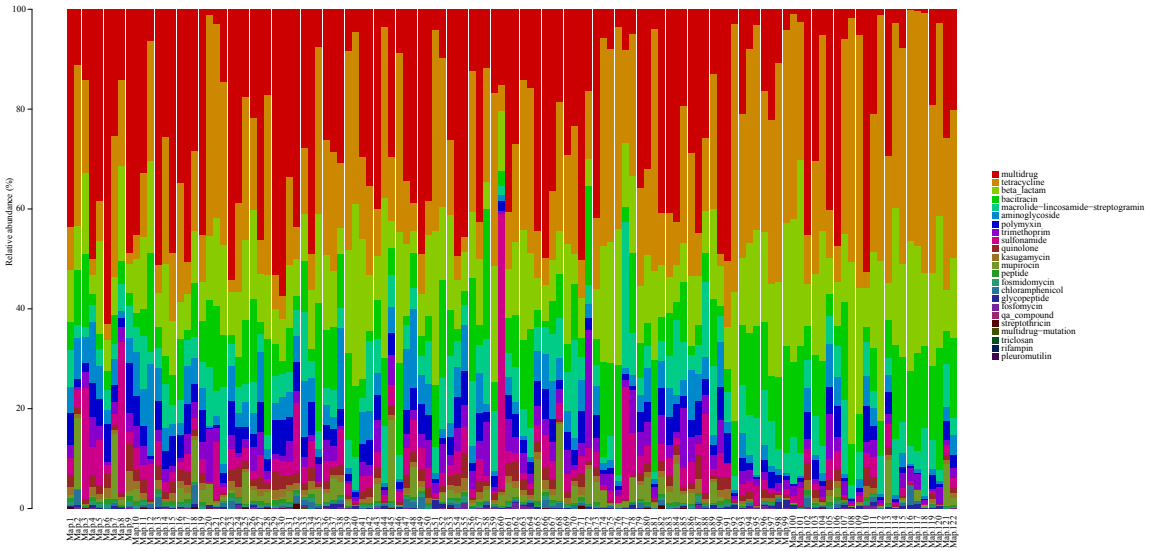


Figure C 1. Relative abundance of different classes of ARGs. deepARG (215) was used to identify and classify ARG-encoding metagenomic reads after searcing them against an ARG database.

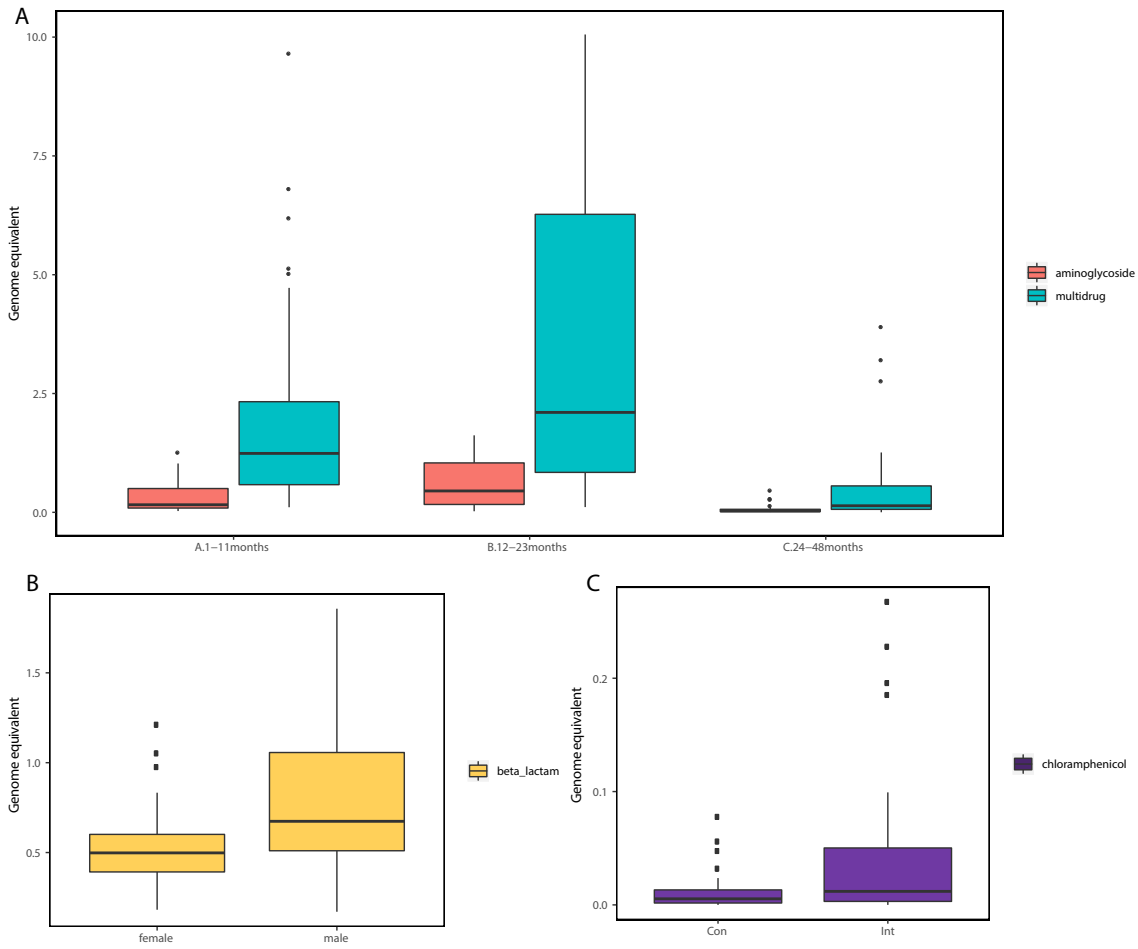


Figure C 2. (A) Higher aminoglycoside and multidrug resistance gene abundance in younger age groups compared to 24-48 months old group. (B) Higher β-lactam resistance gene abundance in male children. (C) Higher chloramphenicol resistance gene abundance in sanitation intervention group compared to control group.

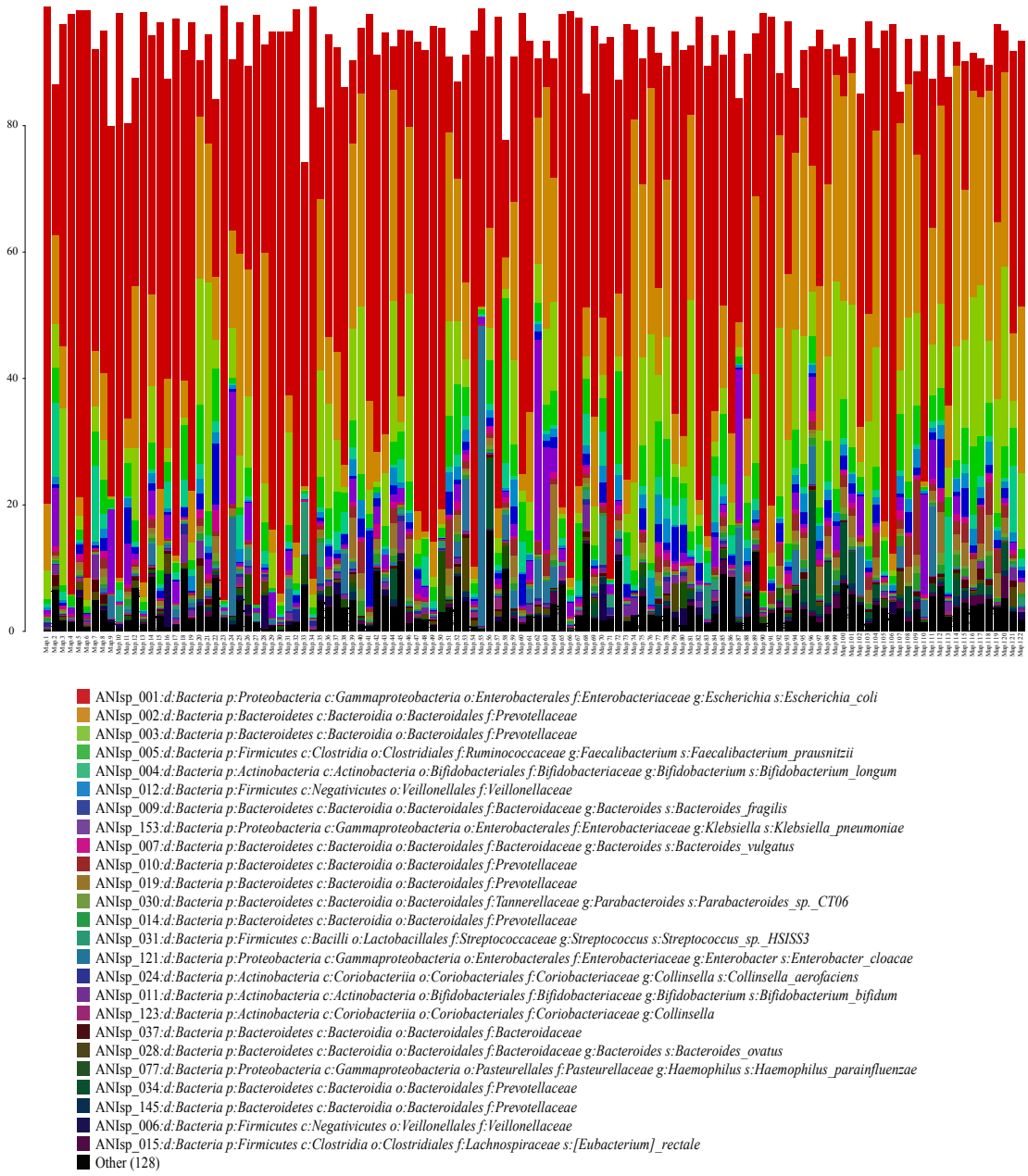


Figure C 3. ARG abundance in the MAGs.

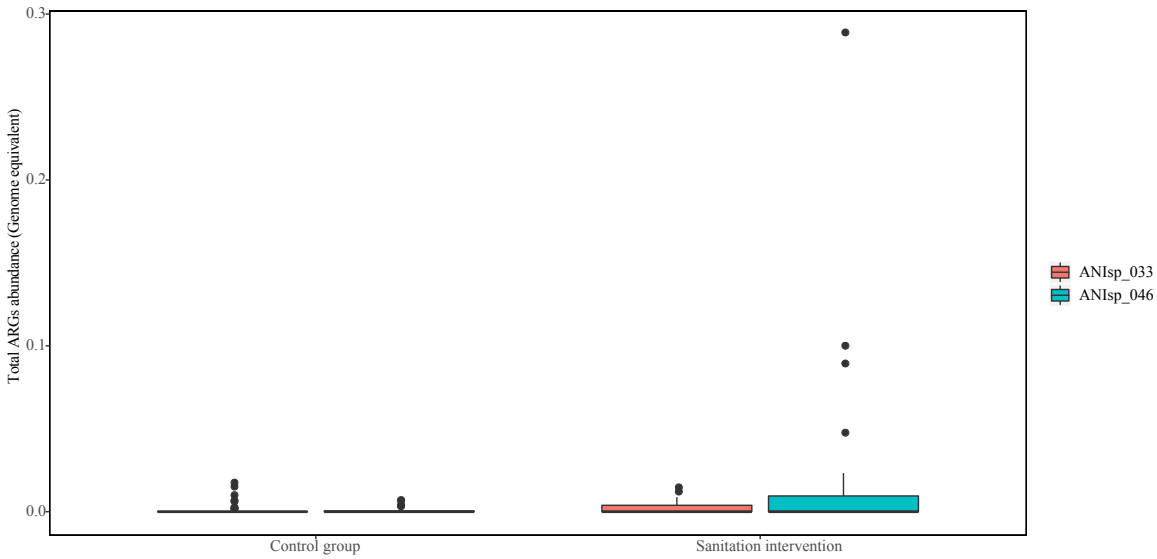


Figure C 4. Two MAGs showing higher abundance of ARGs in the sanitation intervention compared to the control groups.

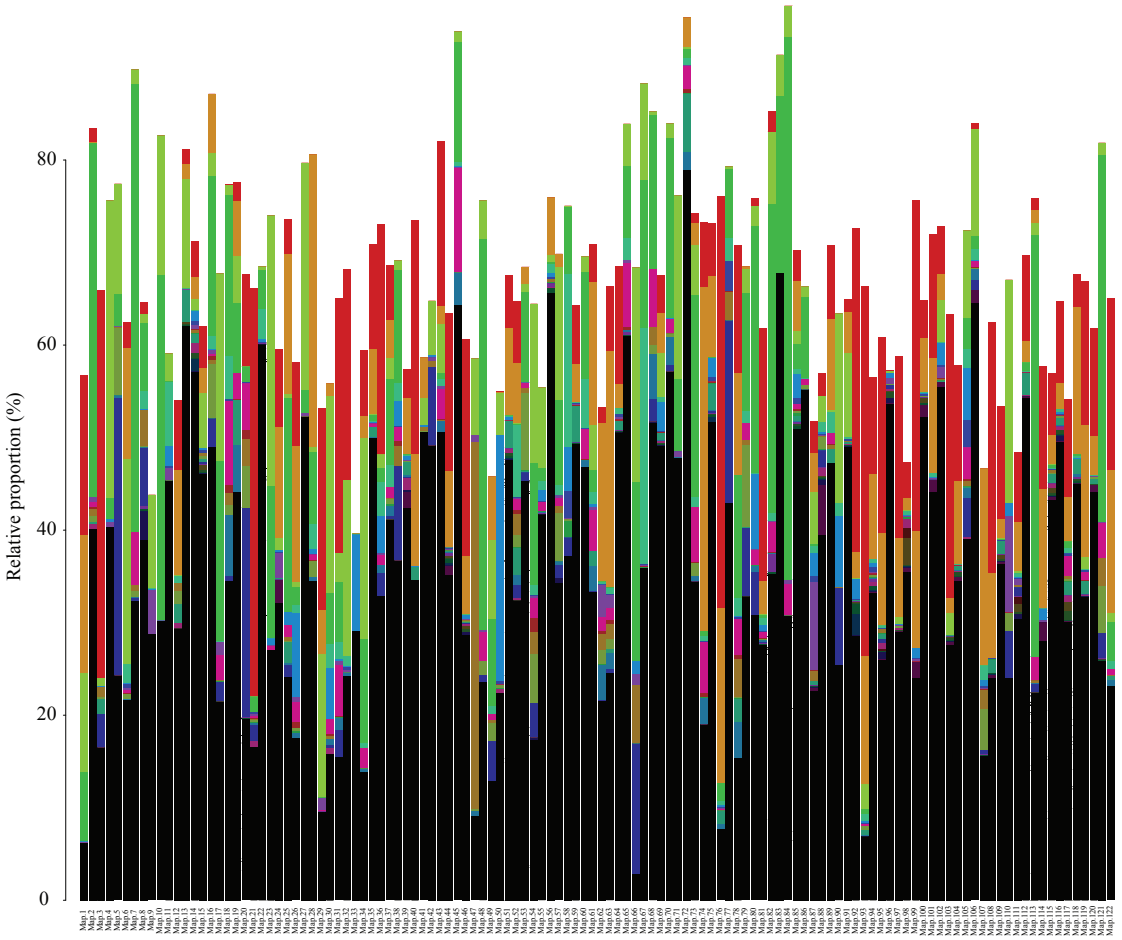


Figure C 5. Relative abundance of MAGs

REFERENCES

1. **Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, Rappuoli R.** 2008. Microbiology in the post-genomic era. Nature Reviews Microbiology **6:**419.

2. **Castelle CJ, Banfield JF.** 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell **172:**1181-1197.

3. **Kim M, Weigand MR, Oh S, Hatt JK, Krishnan R, Tezel U, Pavlostathis SG, Konstantinidis KT.** 2018. Widely used benzalkonium chloride disinfectants can promote antibiotic resistance. Appl. Environ. Microbiol. **84:**e01201-01218.

4. **Mc Cay PH, Ocampo-Sosa AA, Fleming GT.** 2010. Effect of subinhibitory concentrations of benzalkonium chloride on the competitiveness of Pseudomonas aeruginosa grown in continuous culture. Microbiology **156:**30-38.

5. **Schürch A, Arredondo-Alonso S, Willems R, Goering RV.** 2018. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. Clinical Microbiology and Infection **24:**350-354.

6. **Gasc C, Ribière C, Parisot N, Beugnot R, Defois C, Petit-Biderre C, Boucher D, Peyretaillade E, Peyret P.** 2015. Capturing prokaryotic dark matter genomes. Research in microbiology **166:**814-830.

7. **Streit WR, Schmitz RA.** 2004. Metagenomics–the key to the uncultured microbes. Current opinion in microbiology **7:**492-498.

8. **WHO.** 2014. Antimicrobial resistance: global report on surveillance. World Health Organization.

9. **O'Neill J.** 2014. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. Review on Antimicrobial Resistance.

10. **Tezel U, Pavlostathis SG.** 2015. Quaternary ammonium disinfectants: microbial adaptation, degradation and ecology. Current opinion in biotechnology **33:**296-304.

11. **Gerba CP.** 2015. Quaternary ammonium biocides: efficacy in application. Applied and environmental microbiology **81:**464-469.

12. **Cole E, Addison R, Rubino J, Leese K, Dulaney P, Newell M, Wilkins J, Gaber D, Wineinger T, Criger D.** 2003. Investigation of antibiotic and antibacterial agent cross‐resistance in target bacteria from homes of antibacterial product users and nonusers. Journal of Applied microbiology **95:**664-676.

13. **Kücken D, Feucht H-H, Kaulfers P-M.** 2000. Association of *qacE* and *qacE* Δ1 with multiple resistance to antibiotics and antiseptics in clinical isolates of Gram-negative bacteria. FEMS Microbiology Letters **183:**95-98.

14. **Lear J, Maillard J-Y, Dettmar P, Goddard P, Russell A.** 2006. Chloroxylenol- and triclosan-tolerant bacteria from industrial sources—susceptibility to antibiotics and other biocides. International biodeterioration & biodegradation **57:**51-56.

15. **Ortiz S, López P, López V, Martínez-Suárez JV.** 2014. Antibiotic susceptibility in benzalkonium chloride-resistant and-susceptible *Listeria monocytogenes* strains. Foodborne pathogens and disease **11:**517-519.

16. **Anderson RL, Carr JH, Bond WW, Favero MS.** 1997. Susceptibility of vancomycin-resistant enterococci to environmental disinfectants. Infection Control & Hospital Epidemiology **18:**195-199.

17. **Karatzas KA, Randall LP, Webber M, Piddock LJ, Humphrey TJ, Woodward MJ, Coldham NG.** 2008. Phenotypic and proteomic characterization of multiply antibiotic-resistant variants of *Salmonella enterica* Serovar Typhimurium selected following exposure to disinfectants. Applied and environmental microbiology **74:**1508-1516.

18. **Rakic-Martinez M, Drevets DA, Dutta V, Katic V, Kathariou S.** 2011. *Listeria monocytogenes* strains selected on ciprofloxacin or the disinfectant benzalkonium chloride exhibit reduced susceptibility to ciprofloxacin, gentamicin, benzalkonium chloride, and other toxic compounds. Applied and environmental microbiology **77:**8714-8721.

19. **Tandukar M, Oh S, Tezel U, Konstantinidis KT, Pavlostathis SG.** 2013. Long-term exposure to benzalkonium chloride disinfectants results in change of microbial community structure and increased antimicrobial resistance. Environmental science & technology **47:**9730-9738.

20. **Romanova N, Wolffs P, Brovko L, Griffiths M.** 2006. Role of efflux pumps in adaptation and resistance of *Listeria monocytogenes* to benzalkonium chloride. Applied and environmental microbiology **72:**3498-3503.

21. **Condell O, Iversen C, Cooney S, Power KA, Walsh C, Burgess C, Fanning S.** 2012. Efficacy of biocides used in the modern food industry to control *Salmonella*-links between biocide tolerance and resistance to clinically relevant antimicrobial compounds. Applied and environmental microbiology**:**AEM. 07534-07511.

22. **Liu WJ, Fu L, Huang M, Zhang JP, Wu Y, Zhou YS, Zeng J, Wang GX.** 2017. Frequency of antiseptic resistance genes and reduced susceptibility to biocides in carbapenem-resistant Acinetobacter baumannii. Journal of medical microbiology **66:**13-17.

23. **Marple B, Roland P, Benninger M.** 2004. Safety review of benzalkonium chloride used as a preservative in intranasal solutions: an overview of conflicting data and opinions. Otolaryngology-Head and Neck Surgery **130:**131-141.

24. **De Saint Jean M, Brignole F, Bringuier A-F, Bauchet A, Feldmann G, Baudouin C.** 1999. Effects of benzalkonium chloride on growth and survival of Chang conjunctival cells. Investigative ophthalmology & visual science **40:**619-630.

25. **Dixon RE, Kaslow RA, Mackel DC, Fulkerson CC, Mallison GF.** 1976. Aqueous quaternary ammonium antiseptics and disinfectants: use and misuse. JAMA **236:**2415-2417.

26. **Fox J, Beaucage C, Folta C, Thornton G.** 1981. Nosocomial transmission of Serratia marcescens in a veterinary hospital due to contamination by benzalkonium chloride. Journal of clinical microbiology **14:**157-160.

27. **Kaslow RA, Mackel DC, Mallison GF.** 1976. Nosocomial pseudobacteremia: positive blood cultures due to contaminated benzalkonium antiseptic. Jama **236:**2407-2409.

28. **Lee JC, Fialkow PJ.** 1961. Benzalkonium chloride—source of hospital infection with gram-negative bacteria. JAMA **177:**708-710.

29. **Malizia WF, Gangarosa EJ, Goley AF.** 1960. Benzalkonium chloride as a source of infection. New England Journal of Medicine **263:**800-802.

30. **Nakashima A, McCarthy M, Martone W, Anderson R.** 1987. Epidemic septic arthritis caused by Serratia marcescens and associated with a benzalkonium chloride antiseptic. Journal of Clinical Microbiology **25:**1014-1018.

31. **Nasser RM, Rahi AC, Haddad MF, Daoud Z, Irani-Hakime N, Almawi WY.** 2004. Outbreak of Burkholderia cepacia bacteremia traced to contaminated hospital water used for dilution of an alcohol skin antiseptic. Infection Control & Hospital Epidemiology **25:**231-239.

32. **Plotkin SA, Austrian R.** 1958. Bacteremia caused by Pseudomonas sp. following the use of materials stored in solutions of a cationic surface-active agent. The American journal of the medical sciences **235:**621-627.

33. **Sautter RL, Mattman L, Legaspi RC.** 1984. Serratia marcescens meningitis associated with a contaminated benzalkonium chloride solution. Infection Control & Hospital Epidemiology **5:**223-225.

34. **Tiwari TS, Ray B, Jost Jr KC, Rathod MK, Zhang Y, Brown-Elliott BA, Hendricks K, Wallace Jr RJ.** 2003. Forty years of disinfectant failure: outbreak of postinjection Mycobacterium abscessus infection caused by contamination of benzalkonium chloride. Clinical infectious diseases **36:**954-962.

35. **Weber DJ, Rutala WA, Sickbert-Bennett EE.** 2007. Outbreaks associated with contaminated antiseptics and disinfectants. Antimicrobial agents and chemotherapy **51:**4217-4224.

36. **Edgar R, Bibi E.** 1997. MdfA, an Escherichia coli multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. Journal of bacteriology **179:**2274-2280.

37. **Paulsen I, Brown M, Littlejohn T, Mitchell B, Skurray R.** 1996. Multidrug resistance proteins QacA and QacB from Staphylococcus aureus: membrane

topology and identification of residues involved in substrate specificity. Proceedings of the National Academy of Sciences **93:**3630-3635.

38. **Stoitsova SO, Braun Y, Ullrich MS, Weingart H.** 2008. Characterization of the RND-type multidrug efflux pump MexAB-OprM of the plant pathogen Pseudomonas syringae. Appl. Environ. Microbiol. **74:**3387-3393.

39. **Sakagami Y, Yokoyama H, Nishimura H-r, Ose Y, Tashima T.** 1989. Mechanism of resistance to benzalkonium chloride by Pseudomonas aeruginosa. Appl. Environ. Microbiol. **55:**2036-2040.

40. **Mangalappalli-Illathu AK, Vidović S, Korber DR.** 2008. Differential adaptive response and survival of *Salmonella enterica* serovar enteritidis planktonic and biofilm cells exposed to benzalkonium chloride. Antimicrobial agents and chemotherapy **52:**3669-3680.

41. **Oh S, Kurt Z, Tsementzi D, Weigand MR, Kim M, Hatt JK, Tandukar M, Pavlostathis SG, Spain JC, Konstantinidis KT.** 2014. Microbial community degradation of widely used quaternary ammonium disinfectants. Appl. Environ. Microbiol. **80:**5892-5900.

42. **Ertekin E, Konstantinidis KT, Tezel U.** 2016. A Rieske-type oxygenase of Pseudomonas sp. BIOMIG1 converts benzalkonium chlorides to benzyldimethyl amine. Environmental science & technology **51:**175-181.

43. **WHO/UNICEF.** 2017. Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines.

44. **Cairncross S, Blumenthal U, Kolsky P, Moraes L, Tayeh A.** 1996. The public and domestic domains in the transmission of disease. Tropical Medicine & International Health **1:**27-34.

45. **Wagner EG, Lanoix JN, Organization WH.** 1958. Excreta disposal for rural areas and small communities.

46. **Labite H, Lunani I, van der Steen P, Vairavamoorthy K, Drechsel P, Lens P.** 2010. Quantitative Microbial Risk Analysis to evaluate health effects of interventions in the urban water system of Accra, Ghana. Journal of water and health **8:**417-430.

47. **Katukiza A, Ronteltap M, Van Der Steen P, Foppen J, Lens P.** 2014. Quantification of microbial risks to human health caused by waterborne viruses and bacteria in an urban slum. Journal of applied microbiology **116:**447-463.

48. **Knee J, Sumner T, Adriano Z, Berendes D, de Bruijn E, Schmidt W-P, Nalá R, Cumming O, Brown J.** 2018. Risk factors for childhood enteric infection in urban Maputo, Mozambique: A cross-sectional study. PLoS neglected tropical diseases **12:**e0006956.

49. **WHO/UNICEF.** 2015. Progress on sanitation and drinking water: 2015 update and MDG assessment. World Health Organization.

50. **Roche R, Bain R, Cumming O.** 2017. A long way to go–Estimates of combined water, sanitation and hygiene coverage for 25 sub-Saharan African countries. PloS one **12:**e0171783.

51. **Clemente JC, Ursell LK, Parfrey LW, Knight R.** 2012. The impact of the gut microbiota on human health: an integrative view. Cell **148:**1258-1270.

52. **Sonnenburg JL, Xu J, Leip DD, Chen C-H, Westover BP, Weatherford J, Buhler JD, Gordon JI.** 2005. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science **307:**1955-1959.

53. **Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP.** 2012. Human gut microbiome viewed across age and geography. nature **486:**222.

54. **Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, Glickman JN, Siebert R, Baron RM, Kasper DL.** 2012. Microbial exposure during early life has persistent effects on natural killer T cell function. Science **336:**489-493.

55. **Candela M, Perna F, Carnevali P, Vitali B, Ciati R, Gionchetti P, Rizzello F, Campieri M, Brigidi P.** 2008. Interaction of probiotic *Lactobacillus* and *Bifidobacterium* strains with human intestinal epithelial cells: adhesion properties, competition against enteropathogens and modulation of IL-8 production. International journal of food microbiology **125:**286-292.

56. **Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, Tobe T, Clarke JM, Topping DL, Suzuki T.** 2011. Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature **469:**543.

57. **Avershina E, Storrø O, Øien T, Johnsen R, Pope P, Rudi K.** 2014. Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. FEMS microbiology ecology **87:**280-290.

58. **Bergström A, Skov TH, Bahl MI, Roager HM, Christensen LB, Ejlerskov KT, Mølgaard C, Michaelsen KF, Licht TR.** 2014. Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants. Applied and environmental microbiology**:**AEM. 00342-00314.

59. **Cheng J, Ringel-Kulka T, Heikamp-de Jong I, Ringel Y, Carroll I, De Vos WM, Salojärvi J, Satokari R.** 2016. Discordant temporal development of bacterial phyla and the emergence of core in the fecal microbiota of young children. The ISME journal **10:**1002.

60. **Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO.** 2007. Development of the human infant intestinal microbiota. PLoS biology **5:**e177.

61. **Kelly D, King T, Aminov R.** 2007. Importance of microbial colonization of the gut in early life to the development of immunity. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **622:**58-69.

62. **Gottrand F.** 2008. Long-chain polyunsaturated fatty acids influence the immune system of infants. The Journal of nutrition **138:**1807S-1812S.

63. **Bäckhed F, Fraser CM, Ringel Y, Sanders ME, Sartor RB, Sherman PM, Versalovic J, Young V, Finlay BB.** 2012. Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. Cell host & microbe **12:**611-622.

64. **Brown J, Cumming O, Bartram J, Cairncross S, Ensink J, Holcomb D, Knee J, Kolsky P, Liang K, Liang S.** 2015. A controlled, before-and-after trial of an urban sanitation intervention to reduce enteric infections in children: research protocol for the Maputo Sanitation (MapSan) study, Mozambique. BMJ open **5:**e008215.

65. **Humphrey JH.** 2009. Child undernutrition, tropical enteropathy, toilets, and handwashing. The Lancet **374:**1032-1035.

66. **Lunn PG.** 2000. The impact of infection and nutrition on gut function and growth in childhood. Proceedings of the Nutrition Society **59:**147-154.

67. **Ngure FM, Humphrey JH, Mbuya MN, Majo F, Mutasa K, Govha M, Mazarura E, Chasekwa B, Prendergast AJ, Curtis V.** 2013. Formative research on hygiene behaviors and geophagy among infants and young children and implications of exposure to fecal bacteria. The American journal of tropical medicine and hygiene **89:**709-716.

68. **Ngure FM, Reid BM, Humphrey JH, Mbuya MN, Pelto G, Stoltzfus RJ.** 2014. Water, sanitation, and hygiene (WASH), environmental enteropathy, nutrition, and early child development: making the links. Annals of the New York Academy of Sciences **1308:**118-128.

69. **McKay S, Gaudier E, Campbell DI, Prentice AM, Albers R.** 2010. Environmental enteropathy: new targets for nutritional interventions. International Health **2:**172-180.

70. **McCormick B.** 2014, p 23-39. Old Herborn University Seminar Monographs.

71. **Petri WA, Miller M, Binder HJ, Levine MM, Dillingham R, Guerrant RL.** 2008. Enteric infections, diarrhea, and their impact on function and development. The Journal of clinical investigation **118:**1277-1290.

72. **Wulan N, Smith E, Prado E, Apriatni M, Muadz H, Shankar A.** 2015. The long-term impact of water and sanitation on childhood cognition. The FASEB Journal **29:**899.892.

73. **Guerrant RL, DeBoer MD, Moore SR, Scharf RJ, Lima AA.** 2013. The impoverished gut—a triple burden of diarrhoea, stunting and chronic disease. Nature reviews Gastroenterology & hepatology **10:**220.

74. **Abubakar I, Tillmann T, Banerjee A.** 2015. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet **385:**117-171.

75. **Berendes D, Leon J, Kirby A, Clennon J, Raj S, Yakubu H, Robb K, Kartikeyan A, Hemavathy P, Gunasekaran A.** 2017. Household sanitation is associated with lower risk of bacterial and protozoal enteric infections, but not viral infections and diarrhoea, in a cohort study in a low‐income urban neighbourhood in Vellore, India. Tropical Medicine & International Health **22:**1119-1129.

76. **Norman G, Pedley S, Takkouche B.** 2010. Effects of sewerage on diarrhoea and enteric infections: a systematic review and meta-analysis. The Lancet infectious diseases **10:**536-544.

77. **Berendes DM, Kirby AE, Clennon JA, Agbemabiese C, Ampofo JA, Armah GE, Baker KK, Liu P, Reese HE, Robb KA.** 2018. Urban sanitation coverage and environmental fecal contamination: Links between the household and public environments of Accra, Ghana. PloS one **13:**e0199304.

78. **Barreto ML, Genser B, Strina A, Teixeira MG, Assis AMO, Rego RF, Teles CA, Prado MS, Matos SM, Alcântara-Neves NM.** 2010. Impact of a citywide sanitation program in Northeast Brazil on intestinal parasites infection in young children. Environmental health perspectives **118:**1637-1642.

79. **Barreto ML, Genser B, Strina A, Assis AMO, Rego RF, Teles CA, Prado MS, Matos SM, Santos DN, dos Santos LA.** 2007. Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast Brazil: assessment by two cohort studies. The Lancet **370:**1622-1628.

80. **Wolf J, Prüss‐Ustün A, Cumming O, Bartram J, Bonjour S, Cairncross S, Clasen T, Colford Jr JM, Curtis V, De France J.** 2014. Systematic review: assessing the impact of drinking water and sanitation on diarrhoeal disease in low‐and middle‐income settings: systematic review and meta‐regression. Tropical Medicine & International Health **19:**928-942.

81. **Arnold BF, Khush RS, Ramaswamy P, London AG, Rajkumar P, Ramaprabha P, Durairaj N, Hubbard AE, Balakrishnan K, Colford JM.** 2010. Causal inference methods to study nonrandomized, preexisting development interventions. Proceedings of the National Academy of Sciences **107:**22605-22610.

82. **Patil SR, Arnold BF, Salvatore AL, Briceno B, Ganguly S, Colford Jr JM, Gertler PJ.** 2014. The effect of India's total sanitation campaign on defecation behaviors and child health in rural Madhya Pradesh: a cluster randomized controlled trial. PLoS medicine **11:**e1001709.

83. **Clasen T, Boisson S, Routray P, Torondel B, Bell M, Cumming O, Ensink J, Freeman M, Jenkins M, Odagiri M.** 2014. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. The Lancet Global Health **2:**e645-e653.

84. **Pickering AJ, Djebbari H, Lopez C, Coulibaly M, Alzua ML.** 2015. Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial. The Lancet Global Health **3:**e701-e711.

85. **Platts-Mills JA, Babji S, Bodhidatta L, Gratz J, Haque R, Havt A, McCormick BJ, McGrath M, Olortegui MP, Samie A.** 2015. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). The Lancet Global health **3:**e564-e575.

86. **Houpt E, Gratz J, Kosek M, Zaidi AK, Qureshi S, Kang G, Babji S, Mason C, Bodhidatta L, Samie A.** 2014. Microbiologic methods utilized in the MAL-ED cohort study. Clinical Infectious Diseases **59:**S225-S232.

87. **Eibach D, Krumkamp R, Hahn A, Sarpong N, Adu-Sarkodie Y, Leva A, Käsmaier J, Panning M, May J, Tannich E.** 2016. Application of a multiplex PCR assay for the detection of gastrointestinal pathogens in a rural African setting. BMC infectious diseases **16:**150.

88. **Duong VT, Phat VV, Tuyen HT, Dung TTN, Trung PD, Van Minh P, Campbell JI, Le Phuc H, Ha TTT, Ngoc NM.** 2016. Evaluation of Luminex xTAG gastrointestinal pathogen panel assay for detection of multiple diarrheal pathogens in fecal samples in Vietnam. Journal of clinical microbiology **54:**1094-1100.

89. **Katukiza A, Ronteltap M, Niwagaba C, Foppen J, Kansiime F, Lens P.** 2012. Sustainable sanitation technology options for urban slums. Biotechnology advances **30:**964-978.

90. **Bakalian A, Wright A, Otis R, de Azevedo Netto J.** 1994. Simplified sewerage: design guidelines. World Bank Washington, DC.

91. **Paterson C, Mara D, Curtis T.** 2007. Pro-poor sanitation technologies. Geoforum **38:**901-907.

92. **Mara D, Alabaster G.** 2008. A new paradigm for low-cost urban water supplies and sanitation in developing countries. Water policy **10:**119-129.

93. **Tilley E.** 2014. Compendium of sanitation systems and technologies. Eawag.

94. **Giuliano CA, Rybak MJ.** 2015. Efficacy of triclosan as an antimicrobial hand soap and its potential impact on antimicrobial resistance: a focused review. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy **35:**328-336.

95. **Braoudaki M, Hilton A.** 2004. Adaptive resistance to biocides in *Salmonella enterica* and *Escherichia coli* O157 and cross-resistance to antimicrobial agents. Journal of Clinical Microbiology **42:**73-78.

96. **Cottell A, Denyer SP, Hanlon GW, Ochs D, Maillard JY.** 2009. Triclosan-tolerant bacteria: changes in susceptibility to antibiotics. J Hosp Infect **72:**71-76.

97. **Tezel U, Tandukar M, Martinez RJ, Sobecky PA, Pavlostathis SG.** 2012. Aerobic biotransformation of n-tetradecylbenzyldimethylammonium chloride by an enriched *Pseudomonas* spp. community. Environmental science & technology **46:**8714-8722.

98. **Oh S, Kurt Z, Tsementzi D, Weigand MR, Kim M, Hatt JK, Tandukar M, Pavlostathis SG, Spain JC, Konstantinidis KT.** 2014. Microbial community degradation of widely used quaternary ammonium disinfectants. Applied and environmental microbiology **80:**5892-5900.

99. **Çeçen F, Tezel U.** 2017. Hazardous Pollutants in Biological Treatment Systems: Fundamentals and a Guide to Experimental Research. IWA Publishing.

100. **Oh S, Tandukar M, Pavlostathis SG, Chain PS, Konstantinidis KT.** 2013. Microbial community adaptation to quaternary ammonium biocides as revealed by metagenomics. Environmental microbiology **15:**2850-2864.

101. **Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM, Peterson KM.** 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene **166:**175-176.

102. **Mima T, Joshi S, Gomez-Escalada M, Schweizer HP.** 2007. Identification and characterization of TriABC-OpmH, a triclosan efflux pump of *Pseudomonas aeruginosa* requiring two membrane fusion proteins. Journal of bacteriology **189:**7600-7609.

103. **Choi K-H, Kumar A, Schweizer HP.** 2006. A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: application for DNA fragment transfer between chromosomes and plasmid transformation. Journal of microbiological methods **64:**391-397.

104. **WIKLER M.** 2006. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically: approved standard. CLSI (NCCLS) **26:**M7-A7.

105. **Oh S, Caro-Quintero A, Tsementzi D, DeLeon-Rodriguez N, Luo C, Poretsky R, Konstantinidis KT.** 2011. Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. Applied and environmental microbiology **77:**6000-6011.

106. **Ertekin E, Hatt JK, Konstantinidis KT, Tezel U.** 2016. Similar Microbial Consortia and Genes Are Involved in the Biodegradation of Benzalkonium Chlorides in Different Environments. Environmental science & technology **50:**4304-4313.

107. **Tsementzi D, Poretsky R, Rodriguez‐R LM, Luo C, Konstantinidis KT.** 2014. Evaluation of metatranscriptomic protocols and application to the study of freshwater microbial communities. Environmental microbiology reports **6:**640-655.

108. **Cox MP, Peterson DA, Biggs PJ.** 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC bioinformatics **11:**485.

109. **Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM, Konstantinidis KT.** 2011. Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. Proceedings of the National Academy of Sciences **108:**7200-7205.

110. **Noguchi H, Park J, Takagi T.** 2006. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic acids research **34:**5623-5630.

111. **Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R.** 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic acids research **33:**5691-5702.

112. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nature methods **9:**357.

113. **Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF.** 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature **461:**1243-1247.

114. **Deatherage DE, Barrick JE.** 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Engineering and Analyzing Multicellular Systems: Methods and Protocols**:**165-188.

115. **Rodriguez-R LM, Konstantinidis KT.** 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes 2167-9843. PeerJ Preprints.

116. **Loman NJ, Quinlan AR.** 2014. Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics **30:**3399-3401.

117. **Wick RR, Judd LM, Gorrie CL, Holt KE.** 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology **13:**e1005595.

118. **Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK.** 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one **9:**e112963.

119. **Kopylova E, Noé L, Touzet H.** 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics **28:**3211-3217.

120. **Rodriguez-R LM, Konstantinidis KT.** 2014. Estimating coverage in metagenomic data sets and why it matters. The ISME journal **8:**2349.

121. **Liao Y, Smyth GK, Shi W.** 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30:**923-930.

122. **Love MI, Huber W, Anders S.** 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology **15:**1.

123. **Drake JW, Charlesworth B, Charlesworth D, Crow JF.** 1998. Rates of spontaneous mutation. Genetics **148:**1667-1686.

124. **Hardalo C, Edberg SC.** 1997. Pseudomonas aeruginosa: assessment of risk from drinking water. Critical reviews in microbiology **23:**47-75.

125. **Pellett S, Bigley D, Grimes D.** 1983. Distribution of Pseudomonas aeruginosa in a riverine ecosystem. Applied and environmental microbiology **45:**328-332.

126. **Kung VL, Ozer EA, Hauser AR.** 2010. The accessory genome of Pseudomonas aeruginosa. Microbiology and molecular biology reviews **74:**621-641.

127. **Johnson CM, Grossman AD.** 2015. Integrative and conjugative elements (ICEs): what they do and how they work. Annual review of genetics **49:**577-601.

128. **McKinnon P, Davis S.** 2004. Pharmacokinetic and pharmacodynamic issues in the treatment of bacterial infectious diseases. European Journal of Clinical Microbiology and Infectious Diseases **23:**271-288.

129. **Tam VH, Schilling AN, Vo G, Kabbara S, Kwa AL, Wiederhold NP, Lewis RE.** 2005. Pharmacodynamics of polymyxin B against *Pseudomonas aeruginosa*. Antimicrobial agents and chemotherapy **49:**3624-3630.

130. **Burman WJ, Gallicano K, Peloquin C.** 2001. Comparative pharmacokinetics and pharmacodynamics of the rifamycin antibacterials. Clinical pharmacokinetics **40:**327-341.

131. **Chung YJ, Saier MH.** 2002. Overexpression of the Escherichia coli sugE gene confers resistance to a narrow range of quaternary ammonium compounds. Journal of bacteriology **184:**2543-2545.

132. **He G-X, Zhang C, Crow RR, Thorpe C, Chen H, Kumar S, Tsuchiya T, Varela MF.** 2011. SugE, a new member of the SMR family of transporters, contributes to antimicrobial resistance in *Enterobacter cloacae*. Antimicrobial agents and chemotherapy**:**AAC. 00094-00011.

133. **Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S.** 2012. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. Nature genetics **44:**106-110.

134. **Jatsenko T, Tover A, Tegova R, Kivisaar M.** 2010. Molecular characterization of Rif r mutations in *Pseudomonas aeruginosa* and *Pseudomonas putida*. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis **683:**106-114.

135. **Weigand MR, Sundin GW.** 2012. General and inducible hypermutation facilitate parallel adaptation in *Pseudomonas aeruginosa* despite divergent mutation spectra. Proceedings of the National Academy of Sciences **109:**13680-13685.

136. **Moskowitz SM, Ernst RK, Miller SI.** 2004. PmrAB, a two-component regulatory system of *Pseudomonas aeruginosa* that modulates resistance to cationic antimicrobial peptides and addition of aminoarabinose to lipid A. Journal of bacteriology **186:**575-579.

137. **Moskowitz SM, Brannon MK, Dasgupta N, Pier M, Sgambati N, Miller AK, Selgrade SE, Miller SI, Denton M, Conway SP.** 2012. PmrB mutations promote polymyxin resistance of *Pseudomonas aeruginosa* isolated from colistin-treated cystic fibrosis patients. Antimicrobial agents and chemotherapy **56:**1019-1030.

138. **Yan A, Guan Z, Raetz CR.** 2007. An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in *Escherichia coli*. Journal of Biological Chemistry **282:**36077-36089.

139. **Nikaido H, Vaara M.** 1985. Molecular basis of bacterial outer membrane permeability. Microbiological reviews **49:**1.

140. **Vaara M.** 1992. Agents that increase the permeability of the outer membrane. Microbiological reviews **56:**395-411.

141. **Abraham N, Kwon DH.** 2009. A single amino acid substitution in PmrB is associated with polymyxin B resistance in clinical isolate of *Pseudomonas aeruginosa*. FEMS microbiology letters **298:**249-254.

142. **Srikumar R, Kon T, Gotoh N, Poole K.** 1998. Expression of Pseudomonas aeruginosaMultidrug Efflux Pumps MexA-MexB-OprM and MexC-MexD-OprJ in a Multidrug-Sensitive Escherichia coli Strain. Antimicrobial agents and chemotherapy **42:**65-71.

143. **Gotoh N, Tsujimoto H, Tsuda M, Okamoto K, Nomura A, Wada T, Nakahashi M, Nishino T.** 1998. Characterization of the MexC-MexD-OprJ multidrug efflux system in ΔmexA-mexB-oprM mutants of Pseudomonas aeruginosa. Antimicrobial agents and chemotherapy **42:**1938-1943.

144. **Morita Y, Murata T, Mima T, Shiota S, Kuroda T, Mizushima T, Gotoh N, Nishino T, Tsuchiya T.** 2003. Induction of *mexCD-oprJ* operon for a multidrug efflux pump by disinfectants in wild-type *Pseudomonas aeruginosa* PAO1. Journal of Antimicrobial Chemotherapy **51:**991-994.

145. **Poole K, Gotoh N, Tsujimoto H, Zhao Q, Wada A, Yamasaki T, Neshat S, Yamagishi Ji, Li XZ, Nishino T.** 1996. Overexpression of the *mexC–mexD–oprJ* efflux operon in *nfxB*‐type multidrug‐resistant strains of *Pseudomonas aeruginosa*. Molecular microbiology **21:**713-725.

146. **Tiedink J.** 2001. Cationic surfactants: Cationic surfactants in biocides. Handbook of Applied Surface and Colloid Chemistry **1:**318-321.

147. **McDonnell G, Russell AD.** 1999. Antiseptics and disinfectants: activity, action, and resistance. Clinical microbiology reviews **12:**147-179.

148. **Salton M.** 1968. Lytic agents, cell permeability, and monolayer penetrability. The Journal of general physiology **52:**227-252.

149. **Jovović M, Kostić N, Jančić-Stojanović B, Malenović A.** 2012. Investigation of tropicamide and benzalkonium chloride stability using liquid chromatography. Journal of Liquid Chromatography & Related Technologies **35:**231-239.

150. **Li X, Brownawell BJ.** 2010. Quaternary ammonium compounds in urban estuarine sediment environments-a class of contaminants in need of increased attention? Environmental science & technology **44:**7561-7568.

151. **Martínez-Carballo E, González-Barreiro C, Sitka A, Kreuzinger N, Scharf S, Gans O.** 2007. Determination of selected quaternary ammonium compounds by liquid chromatography with mass spectrometry. Part II. Application to sediment and sludge samples in Austria. Environmental Pollution **146:**543-547.

152. **Martínez-Carballo E, Sitka A, González-Barreiro C, Kreuzinger N, Fürhacker M, Scharf S, Gans O.** 2007. Determination of selected quaternary ammonium compounds by liquid chromatography with mass spectrometry. Part I. Application to surface, waste and indirect discharge water samples in Austria. Environmental Pollution **145:**489-496.

153. **Kümmerer K, Eitel A, Braun U, Hubner P, Daschner F, Mascart G, Milandri M, Reinthaler F, Verhoef J.** 1997. Analysis of benzalkonium chloride in the effluent from European hospitals by solid-phase extraction and high-performance liquid chromatography with post-column ion-pairing and fluorescence detection. Journal of chromatography A **774:**281-286.

154. **Rucker RR, Johnson HE, Ordal EJ.** 1949. An investigation of the bactericidal action and fish toxicity of two homologous series of quaternary ammonium compounds. Journal of bacteriology **57:**225.

155. **D'Arcy P, Taylor E.** 1962. Quaternary ammonium compounds in medicinal chemistry. I. Journal of Pharmacy and Pharmacology **14:**129-146.

156. **Tiwari TS, Ray B, Jost KC, Rathod MK, Zhang Y, Brown-Elliott BA, Hendricks K, Wallace RJ.** 2003. Forty years of disinfectant failure: outbreak of postinjection *Mycobacterium abscessus* infection caused by contamination of benzalkonium chloride. Clinical infectious diseases **36:**954-962.

157. **Ahn Y, Kim JM, Kweon O, Kim S-J, Jones RC, Woodling K, da Costa GG, LiPuma JJ, Hussong D, Marasa BS.** 2016. Intrinsic Resistance of *Burkholderia cepacia* Complex to Benzalkonium Chloride. mBio **7:**e01716-01716.

158. **Nishino K, Yamaguchi A.** 2001. Analysis of a complete library of putative drug transporter genes in Escherichia coli. Journal of bacteriology **183:**5803-5812.

159. **He G-X, Kuroda T, Mima T, Morita Y, Mizushima T, Tsuchiya T.** 2004. An H+-coupled multidrug efflux pump, PmpM, a member of the MATE family of transporters, from *Pseudomonas aeruginosa*. Journal of bacteriology **186:**262-265.

160. **Stoitsova SO, Braun Y, Ullrich MS, Weingart H.** 2008. Characterization of the RND-type multidrug efflux pump MexAB-OprM of the plant pathogen *Pseudomonas syringae*. Applied and environmental microbiology **74:**3387-3393.

161. **Paulsen I, Littlejohn T, Rådström P, Sundström L, Sköld O, Swedberg G, Skurray R.** 1993. The 3'conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. Antimicrobial Agents and Chemotherapy **37:**761-768.

162. **Heir E, Sundheim G, Holck AL.** 1998. The *Staphylococcus qacH* gene product: a new member of the SMR family encoding multidrug resistance. FEMS microbiology letters **163:**49-56.

163. **Nagai K, Murata T, Ohta S, Zenda H, Ohnishi M, Hayashi T.** 2003. Two different mechanisms are involved in the extremely high‐level benzalkonium chloride resistance of a *Pseudomonas fluorescens* strain. Microbiology and immunology **47:**709-715.

164. **Braoudaki M, Hilton AC.** 2005. Mechanisms of resistance in *Salmonella enterica* adapted to erythromycin, benzalkonium chloride and triclosan. International journal of antimicrobial agents **25:**31-37.

165. **Kahm M, Hasenbrink G, Lichtenberg-Fraté H, Ludwig J, Kschischo M.** 2010. grofit: fitting biological growth curves with R. Journal of Statistical Software **33:**1-21.

166. **Schuster M, Lostroh CP, Ogi T, Greenberg E.** 2003. Identification, timing, and signal specificity of *Pseudomonas aeruginosa* quorum-controlled genes: a transcriptome analysis. Journal of bacteriology **185:**2066-2079.

167. **Wagner VE, Bushnell D, Passador L, Brooks AI, Iglewski BH.** 2003. Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment. Journal of bacteriology **185:**2080-2095.

168. **Guo W, Cui S, Xu X, Wang H.** 2014. Resistant mechanism study of benzalkonium chloride selected *Salmonella* Typhimurium mutants. Microbial Drug Resistance **20:**11-16.

169. **Heo A, Jang H-J, Sung J-S, Park W.** 2014. Global transcriptome and physiological responses of *Acinetobacter oleivorans* DR1 exposed to distinct classes of antibiotics. PLoS One **9:**e110215.

170. **POOLE K, HANCOCK RE.** 1984. Phosphate transport in *Pseudomonas aeruginosa*. European Journal of Biochemistry **144:**607-612.

171. **Nikata T, Sakai Y, Shibata K, Kato J, Kuroda A, Ohtake H.** 1996. Molecular analysis of the phosphate-specific transport (pst) operon of *Pseudomonas aeruginosa*. Molecular and General Genetics MGG **250:**692-698.

172. **Martinez A, Tyson GW, DeLong EF.** 2010. Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. Environmental microbiology **12:**222-238.

173. **Quinn JP, Kulakova AN, Cooley NA, McGrath JW.** 2007. New ways to break an old bond: the bacterial carbon–phosphorus hydrolases and their role in biogeochemical phosphorus cycling. Environmental microbiology **9:**2392-2400.

174. **White AK, Metcalf WW.** 2007. Microbial metabolism of reduced phosphorus compounds. Annu. Rev. Microbiol. **61:**379-400.

175. **Comolli JC, Donohue TJ.** 2002. *Pseudomonas aeruginosa* RoxR, a response regulator related to *Rhodobacter sphaeroides* PrrA, activates expression of the cyanide‐insensitive terminal oxidase. Molecular microbiology **45:**755-768.

176. **Claudi B, Spröte P, Chirkova A, Personnic N, Zankl J, Schürmann N, Schmidt A, Bumann D.** 2014. Phenotypic variation of *Salmonella* in host tissues delays eradication by antimicrobial chemotherapy. Cell **158:**722-733.

177. **Brauner A, Fridman O, Gefen O, Balaban NQ.** 2016. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. Nature Reviews Microbiology **14:**320-330.

178. **Nakata K, Tsuchido T, Matsumura Y.** 2011. Antimicrobial cationic surfactant, cetyltrimethylammonium bromide, induces superoxide stress in *Escherichia coli* cells. Journal of applied microbiology **110:**568-579.

179. **Ceragioli M, Mols M, Moezelaar R, Ghelardi E, Senesi S, Abee T.** 2010. Comparative transcriptomic and phenotypic analysis of the responses of *Bacillus cereus* to various disinfectant treatments. Applied and environmental microbiology **76:**3352-3360.

180. **Mangalappalli-Illathu AK, Korber DR.** 2006. Adaptive resistance and differential protein expression of *Salmonella enterica* serovar Enteritidis biofilms exposed to benzalkonium chloride. Antimicrobial agents and chemotherapy **50:**3588-3596.

181. **Kvint K, Nachin L, Diez A, Nyström T.** 2003. The bacterial universal stress protein: function and regulation. Current opinion in microbiology **6:**140-145.

182. **Lee J, Hiibel S, Reardon K, Wood T.** 2010. Identification of stress‐related proteins in *Escherichia coli* using the pollutant cis‐dichloroethylene. Journal of applied microbiology **108:**2088-2102.

183. **Zheng M, Wang X, Templeton LJ, Smulski DR, LaRossa RA, Storz G.** 2001. DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. Journal of bacteriology **183:**4562-4570.

184. **Lewenza S.** 2013. Extracellular DNA-induced antimicrobial peptide resistance mechanisms in *Pseudomonas aeruginosa*. Frontiers in microbiology **4:**21.

185. **Johnson L, Mulcahy H, Kanevets U, Shi Y, Lewenza S.** 2012. Surface-localized spermidine protects the *Pseudomonas aeruginosa* outer membrane from antibiotic treatment and oxidative stress. Journal of bacteriology **194:**813-826.

186. **Mima T, Kohira N, Li Y, Sekiya H, Ogawa W, Kuroda T, Tsuchiya T.** 2009. Gene cloning and characteristics of the RND-type multidrug efflux pump MuxABC-OpmB possessing two RND components in *Pseudomonas aeruginosa*. Microbiology **155:**3509-3517.

187. **Yang L, Chen L, Shen L, Surette M, Duan K.** 2011. Inactivation of MuxABC-OpmB transporter system in *Pseudomonas aeruginosa* leads to increased ampicillin

and carbenicillin resistance and decreased virulence. The Journal of Microbiology **49:**107-114.

188. **Chiang W-C, Pamp SJ, Nilsson M, Givskov M, Tolker-Nielsen T.** 2012. The metabolically active subpopulation in *Pseudomonas aeruginosa* biofilms survives exposure to membrane-targeting antimicrobials via distinct molecular mechanisms. FEMS Immunology & Medical Microbiology **65:**245-256.

189. **Gensberg K, Smith AW, Brinkman FS, Hancock RE.** 1999. Identification of oprG, a gene encoding a major outer membrane protein of *Pseudomonas aeruginosa*. Journal of Antimicrobial Chemotherapy **43:**607-608.

190. **Peng X, Xu C, Ren H, Lin X, Wu L, Wang S.** 2005. Proteomic Analysis of the Sarcosine-Insoluble Outer Membrane Fraction of *Pseudomonas a eruginosa* Responding to Ampicilin, Kanamycin, and Tetracycline Resistance. Journal of proteome research **4:**2257-2265.

191. **McPhee JB, Tamber S, Bains M, Maier E, Gellatly S, Lo A, Benz R, Hancock RE.** 2009. The major outer membrane protein OprG of *Pseudomonas aeruginosa* contributes to cytotoxicity and forms an anaerobically regulated, cation-selective channel. FEMS microbiology letters **296:**241-247.

192. **Touw DS, Patel DR, Van Den Berg B.** 2010. The crystal structure of OprG from *Pseudomonas aeruginosa*, a potential channel for transport of hydrophobic molecules across the outer membrane. PLoS One **5:**e15016.

193. **Cowley NL, Forbes S, Amézquita A, McClure P, Humphreys GJ, McBain AJ.** 2015. Effects of formulation on microbicide potency and mitigation of the development of bacterial insusceptibility. Applied and environmental microbiology **81:**7330-7338.

194. **Ismail ZZ, Tezel U, Pavlostathis SG.** 2010. Sorption of quaternary ammonium compounds to municipal sludge. Water research **44:**2303-2313.

195. **Troeger C, Blacker BF, Khalil IA, Rao PC, Cao S, Zimsen SR, Albertson SB, Stanaway JD, Deshpande A, Abebe Z.** 2018. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Infectious Diseases **18:**1211-1228.

196. **Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF.** 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. The Lancet **382:**209-222.

197. **Platts-Mills JA, McCormick BJ, Kosek M, Pan WK, Checkley W, Houpt ER.** 2014. Methods of analysis of enteropathogen infection in the MAL-ED cohort study. Clinical Infectious Diseases **59:**S233-S238.

198. **Investigators TM-EN.** 2014. The MAL-ED study: a multinational and multidisciplinary approach to understand the relationship between enteric pathogens, malnutrition, gut physiology, physical growth, cognitive development, and immune responses in infants and children up to 2 years of age in resource-poor environments. Clinical Infectious Diseases **59:**S193-S206.

199. **Mbuya MN, Humphrey JH.** 2016. Preventing environmental enteric dysfunction through improved water, sanitation and hygiene: an opportunity for stunting reduction in developing countries. Maternal & child nutrition **12:**106-120.

200. **Clasen TF, Bostoen K, Schmidt WP, Boisson S, Fung ICH, Jenkins MW, Scott B, Sugden S, Cairncross S.** 2010. Interventions to improve disposal of human excreta for preventing diarrhoea. Cochrane Database of Systematic Reviews.

201. **Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA.** 2008. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. Bmj **336:**601-605.

202. **Arnold BF, Galiani S, Ram PK, Hubbard AE, Briceño B, Gertler PJ, Colford Jr JM.** 2013. Optimal recall period for caregiver-reported illness in risk factor and intervention studies: a multicountry study. American journal of epidemiology **177:**361-370.

203. **Feikin DR, Audi A, Olack B, Bigogo GM, Polyak C, Burke H, Williamson J, Breiman RF.** 2010. Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. International journal of epidemiology **39:**450-458.

204. **Zafar SN, Luby S, Mendoza C.** 2010. Recall errors in a weekly survey of diarrhoea in Guatemala: determining the optimal length of recall. Epidemiology & Infection **138:**264-269.

205. **Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L.** 2011. Urbanisation and infectious diseases in a globalised world. The Lancet infectious diseases **11:**131-141.

206. **Nakamura S, Maeda N, Miron IM, Yoh M, Izutsu K, Kataoka C, Honda T, Yasunaga T, Nakaya T, Kawai J.** 2008. Metagenomic diagnosis of bacterial infections. Emerging infectious diseases **14:**1784.

207. **Doan T, Acharya NR, Pinsky BA, Sahoo MK, Chow ED, Banaei N, Budvytiene I, Cevallos V, Zhong L, Zhou Z.** 2017. Metagenomic DNA Sequencing for the Diagnosis of Intraocular Infections. Ophthalmology **124:**1247-1248.

208. **Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P.** 2019. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell.

209. **De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P.** 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proceedings of the National Academy of Sciences **107:**14691-14696.

210. **Baqui AH, Black RE, Yunus M, Hoque AA, Chowdhury H, Sack RB.** 1991. Methodological issues in diarrhoeal diseases epidemiology: definition of diarrhoeal episodes. International journal of epidemiology **20:**1057-1063.

211. **Rotmistrovsky K, Agarwala R.** 2011. BMTagger: Best Match Tagger for removing human reads from metagenomics datasets.

212. **Rodriguez-r LM, Konstantinidis KT.** 2013. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics **30:**629-635.

213. **Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM.** 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome biology **17:**132.

214. **Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N.** 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nature methods **12:**902.

215. **Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L.** 2018. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome **6:**23.

216. **Chen L, Zheng D, Liu B, Yang J, Jin Q.** 2015. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. Nucleic acids research **44:**D694-D697.

217. **Orellana LH, Rodriguez-R LM, Konstantinidis KT.** 2016. ROCker: accurate detection and quantification of target genes in short-read metagenomic data sets by modeling sliding-window bitscores. Nucleic acids research **45:**e14-e14.

218. **Zhang J, Kobert K, Flouri T, Stamatakis A.** 2013. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics **30:**614-620.

219. **Peng Y, Leung HC, Yiu S-M, Chin FY.** 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics **28:**1420-1428.

220. **Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW.** 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome **2:**26.

221. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW.** 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome research **25:**1043-1055.

222. **Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT.** 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic acids research.

223. **Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H.** 2015. Package 'vegan'. Community ecology package, version **2**.

224. **Ogle DH.** 2016. Introductory fisheries analyses with R. Chapman and Hall/CRC.

225. **Pena-Gonzalez A, Rodriguez-R LM, Marston CK, Gee JE, Gulvik CA, Kolton CB, Saile E, Frace M, Hoffmaster AR, Konstantinidis KT.** 2018. Genomic Characterization and Copy Number Variation of Bacillus anthracis Plasmids pXO1 and pXO2 in a Historical Collection of 412 Strains. MSystems **3:**e00065-00018.

226. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J.** 2011. Fast, scalable generation of high‐quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology **7:**539.

227. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one **5:**e9490.

228. **Letunic I, Bork P.** 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research **44:**W242-W245.

229. **Grzeskowiak L, Collado MC, Mangani C, Maleta K, Laitinen K, Ashorn P, Isolauri E, Salminen S.** 2012. Distinct gut microbiota in southeastern African and northern European infants. Journal of pediatric gastroenterology and nutrition **54:**812-816.

230. **Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, Finn RD.** 2019. A new genomic blueprint of the human gut microbiota. Nature**:**1.

231. **Center for Disease Dynamics EP.** 2015. Global Antibiotic Resistance Partnership-Mozambique Working Group. 2015. Situation Analysis and Recommendations: Antibiotic Use and Resistance in Mozambique.

232.**Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI.** 2003. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science **299:**2074-2076.

233. **Stappenbeck TS, Hooper LV, Gordon JI.** 2002. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. Proceedings of the National Academy of Sciences **99:**15451-15455.

234. **Hooper LV, Xu J, Falk PG, Midtvedt T, Gordon JI.** 1999. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. Proceedings of the National Academy of Sciences **96:**9833-9838.

235.**Curtis MM, Hu Z, Klimko C, Narayanan S, Deberardinis R, Sperandio V.** 2014. The gut commensal Bacteroides thetaiotaomicron exacerbates enteric infection through modification of the metabolic landscape. Cell host & microbe **16:**759-769.

236. **Bhatti MA, Frank MO.** 2000. Veillonella parvula meningitis: case report and review of Veillonella infections. Clinical Infectious Diseases **31:**839-840.

237.**Africa C, Nel J, Stemmet M.** 2014. Anaerobes and bacterial vaginosis in pregnancy: virulence factors contributing to vaginal colonisation. International journal of environmental research and public health **11:**6979-7000.

238. **Lerche M, Reuter G.** 1960. Contribution to the technique of isolation and differentiation of the aerobic" Lactobacilli"(Genus Beijerinck). Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene, Abt. I (Originale) **179:**354-370.

239. **SHARPE ME, Latham M, GARVIE EI, Zirngibl J, Kandler O.** 1973. Two New Species of Lactobacillus Isolated from the Bovine Rumen, Lactobacillus ruminis sp. nov. and Lactobacillus vitulinus sp. nov. Microbiology **77:**37-49.

240. **Reuter G.** 2001. The Lactobacillus and Bifidobacterium microflora of the human intestine: composition and succession. Current issues in intestinal microbiology **2:**43-53.

241. **Tannock G, Munro K, Harmsen H, Welling G, Smart J, Gopal P.** 2000. Analysis of the fecal microflora of human subjects consuming a probiotic product containing Lactobacillus rhamnosusDR20. Appl. Environ. Microbiol. **66:**2578-2588.

242. **Taweechotipatr M, Iyer C, Spinler JK, Versalovic J, Tumwasorn S.** 2009. Lactobacillus saerimneri and Lactobacillus ruminis: novel human-derived probiotic strains with immunomodulatory activities. FEMS microbiology letters **293:**65-72.

243. **Jiménez E, Marín ML, Martín R, Odriozola JM, Olivares M, Xaus J, Fernández L, Rodríguez JM.** 2008. Is meconium from healthy newborns actually sterile? Research in microbiology **159:**187-193.

244. **Nowrouzian F, Hesselmar B, Saalman R, Strannegård I-L, Åberg N, Wold AE, Adlerberth I.** 2003. Escherichia coli in infants' intestinal microflora: colonization rate, strain turnover, and virulence gene carriage. Pediatric research **54:**8.

245. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic escherichia coli. Nature reviews microbiology **2:**123.

246. **Aijuka M, Santiago AE, Girón JA, Nataro JP, Buys EM.** 2018. Enteroaggregative Escherichia coli is the predominant diarrheagenic E. coli pathotype among irrigation water and food sources in South Africa. International Journal of Food Microbiology **278:**44-51.

247. **Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY.** 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet **380:**2095-2128.

248. **Grindley ND, Whiteson KL, Rice PA.** 2006. Mechanisms of site-specific recombination. Annu. Rev. Biochem. **75:**567-605.

249. **Hickman AB, Chandler M, Dyda F.** 2010. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. Critical reviews in biochemistry and molecular biology **45:**50-69.

250. **Lee CA, Babic A, Grossman AD.** 2010. Autonomous plasmid‐like replication of a conjugative transposon. Molecular microbiology **75:**268-279.